DOCUMENT RESUME

ED 402 336 TM 025 915

AUTHOR DuBois, David A.; And Others
TITLE A Cognitively-Oriented Approach to Task Analysis and Test Development.
INSTITUTION Personnel Decisions Research Institutes, Inc., Minneapolis, MN.
SPONS AGENCY Office of Naval Research, Arlington, Va.
PUB DATE 31 Dec 95
CONTRACT N00014-93-C-0199
NOTE 74p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Cognitive Processes; *Job Analysis; *Personnel Evaluation; Protocol Analysis; *Task Analysis; *Test Construction; Test Items; Test Validity
IDENTIFIERS *Expertise

ABSTRACT

Clear descriptions of job expertise are required to support applications and improvements in personnel training and job performance. This report describes a practical approach to task analysis that integrates the issues, content, and methods of cognitive science and personnel psychology. Cognitively oriented task analysis employs a breadth, then depth, strategy for identifying job expertise. Starting with a task-by-knowledge framework, job expertise is successively elaborated using interviews, expert ratings, and protocol analyses. The application of task analysis results to the development of written performance measures is described to illustrate the contributions of this approach to measurement validity. Task analysis results show that much of what has been missing in using existing task analysis methods is the mental aspects of performance related to interactions among task dimensions, task characteristics, and contexts. Two appendixes provide an example of knowledge elicitation and representation and item writing guidelines for performance measures. (Contains 2 figures, 17 tables, 16 references, 1 appendix table, 5 appendix figures, and 19 appendix references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 402 336

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

A Cognitively-Oriented Approach to Task Analysis and Test Development

David A. DuBois
Psychological Systems & Research, Inc.
1975 Willow Ridge Circle
Kent, Ohio 42440

Valerie L. Shalin
State University of New York
Department of Industrial Engineering
Buffalo, New York 14260

Keith R. Levi
Maharishi University of Management
Fairfield, Iowa 52557

Walter C. Borman
Personnel Decisions Research Institutes, Inc., and
University of South Florida
Department of Psychology
4202 Fowler Avenue
Tampa, Florida 33620

This research was supported by the Office of Naval Research, Cognitive Science Program, and Manpower, Personnel and Training Program, under Contract No. N00014-93-C-0199. Reproduction in whole or part is permitted for any purpose of the United States Government.

Approved for public release; distribution unlimited.

Personnel Decisions Research Institutes, Inc.
43 Main Street SE, Suite 405
Minneapolis, MN 55414
(612) 331-3680

TM025905

BEST COPY AVAILABLE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	31 December 1995	Final, 5 May 1993 to 31 December 1995

4. TITLE AND SUBTITLE

A Cognitively-Oriented Approach to Task Analysis and Test Design

5. FUNDING NUMBERS

N00014-93-C-0199

6. AUTHOR(S)

David A. DuBois, Valerie L. Shalin
Keith R. Levi, Walter C. Borman

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Personnel Decisions Research Institutes, Inc.
43 Main Street SE
Riverplace, Suite 405
Minneapolis, MN 55414

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research
ONR 251B RD
Ballston Tower One
800 N. Quincy
Arlington, VA 22217-5660

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Clear descriptions of job expertise are required to support applications and improvements in personnel training and job performance. This report describes a practical approach to task analysis which integrates the issues, content, and methods of cognitive science and personnel psychology. Cognitively-oriented task analysis employs a breadth then depth strategy for identifying job expertise. Starting with a task by knowledge framework, job expertise is successively elaborated using interviews, expert ratings, and protocol analyses. The application of task analysis results to the development of written performance measures are described to illustrate the contributions of this approach to measurement validity.

14. SUBJECT TERMS

Task analysis, test design, performance measurement, job knowledge, expertise

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89)
Prescribed by ANSI Std 239-18 298-102

BEST COPY AVAILABLE

Table of Contents

Introduction 1
Intended Audience 1
Organization of this Report 1

Section 1: Describing Job Expertise 2
Cognitively-Oriented Task Analyses 2
General Features 2
Model-Based Approach 2
Representative Sampling 2
Cognitive Focus 3
Work Performance in Context 3
The Nature of Job Performance 3
A Model of Job Expertise 4

Section 2: Description of Cognitively-Oriented Task Analysis Methods 9
Step 1: Plan the Project 10
Step 2: Analyze Tasks 11
Step 3: Identify Diagnostic Tasks 18
Step 4: Elicit Detailed Job Knowledge 21
Step 5: Represent Job Expertise 25
Summary of Cognitively-Oriented Task Analysis 29

Section 3: Developing Written Performance Measures 32
Specifying Test Content 32
Selecting Questions to Write 34
Writing Items 37
Question Stems 38
Response Alternatives 40
Reliability and Validity 41
Conclusions 44

References 45

Appendix A: An Example of Knowledge Elicitation and Representation 47

Appendix B: Item Writing Guidelines for Written Performance Measures 60

Figures

1. A completed performance example form for the computer technician job 17
2. Portions of a plan-goal graph for the computer technician job 28

Tables

1. A Task By Knowledge Framework of Job Expertise 4
2. Knowledge Requirements For Performance 8
3. Cognitively-Oriented Task Analysis 10
4. Effects of Task Characteristics on Knowledge Requirements 12
5. Organization of a Job Analysis Interview 13
6. A Guide for Interview Probes 14
7. Description of Expertise for Computer Technicians 19
8. A Partial List of Land Navigation Tasks 26
9. Criteria for Selecting Task Analysis Methods 30
10. A Comparison of Task Analysis Methods 31
11. Test Development Process 32
12. Description of Land Navigation Expertise 33
13. Portion of a Method by Knowledge Element Matrix 35
14. Detailed Test Specifications 36
15. Reliability and Validity of Expert Judgments 41
16. Content Analysis of Land Navigation Tests 42
17. Correlations of Job Knowledge and Performance Measures 43

A Cognitively-Oriented Approach to Task Analysis and Test Development

David A. DuBois, Valerie L. Shalin, Keith R. Levi, and Walter C. Borman

Introduction

This report describes the workplace application of cognitive methods to task analysis and test development. Task analyses are essential to improving personnel performance, including the development of effective programs for selecting, training, and managing performance. Traditionally, task analyses have focused systematically on describing the *behavior* of competent performers. Consequently, measures for predicting, evaluating, or diagnosing performance have also emphasized the behavioral content of performance.

Alternatively, cognitive methods hold considerable promise for improvements in personnel training and performance by revealing the *thought processes* experts use to achieve superior performance. Cognitive methods extend traditional approaches that describe *what* tasks get performed by identifying *how* these tasks are done. This involves describing the critical cognitive content and processes that underlie observable behaviors. The mental aspects of behavior--the goals, strategies, decisions, and prior knowledge--indicate unique and important job content relevant to training, testing, and performance.

Achieving an optimal balance between quality and cost is a traditional challenge for task analyses employed in support of practical applications. We found it necessary to incorporate task analysis methods from both behavior-based and cognitive-focused approaches to thoroughly and practically describe job expertise. Based on personnel psychology, behavior-based methods address the breadth of tasks performed in the workplace. Methods from cognitive science effectively describe the depth of knowledge employed during task performance. The two approaches complement each other well. Hence, we label our approach 'cognitively-oriented task analyses' to recognize the contributions of both. By integrating both approaches, the nature of job expertise can be identified systematically and in a cost effective manner. This report describes the methods employed in cognitively-oriented task analysis, illustrates their use with examples, and discusses the application of this task analysis approach to the development of performance measures.

Intended Audience

The intended audiences for this report are persons responsible for developing human resource (HR) applications such as training objectives and curricula, performance aids (e.g., intelligent tutors) and performance measures. In the military services, these people are often job experts serving as instructors, curriculum designers, and test developers. This report is written for these job experts to assist them in completing their instructional goals. It may also be useful to researchers interested in applying cognitive science to workplace applications.

Organization of this Report

This report is organized into three sections. We begin by first presenting some distinguishing features of our task analysis approach and by describing a general model of job expertise. The second section describes the methods employed in cognitively-oriented task analysis. In the third section, we discuss how results from these methods can be employed to improve the development of performance tests. In Appendix A, we illustrate our knowledge elicitation approach using protocols obtained from our work with computer technicians. We provide some guidelines for developing written performance measures in Appendix B.

Section 1: Describing Job Expertise

Cognitively-Oriented Task Analyses

Cognitively-oriented task analysis involves three phases: description of tasks performed, identification of diagnostic tasks, and elicitation of knowledge that supports task performance. We incorporate techniques from personnel psychology to identify the tasks that comprise a job and to target the more resource-intensive cognitive methods to the most relevant tasks. We utilize cognitive methods to elicit in detail the knowledge requirements of performance.

This breadth-then-depth strategy takes advantage of the complementary nature of task analysis methods employed by personnel psychology and cognitive science. Personnel psychology procedures are task-focused and more cost effective, but suffer from biases and omissions inherent in retrospective self-report methods. Cognitive science methods provide contextually rich, detailed accounts of job knowledge but are very resource intensive to use. Hence, we adapt procedures from personnel psychology to describe job tasks, then target procedures from cognitive science to those tasks that are most informative of job expertise.

In addition to their individual contributions, combining the two approaches to task analysis also yields new insights into the nature of job expertise. In particular, the unique contribution of this cognitively-oriented approach results from identifying tasks and knowledge, essential to competent performance, that were previously implicit. We applied this approach to the computer technician's job and Marine land navigation performance to develop written performance measures (DuBois & Shalin, 1995). Based on our results, this cognitively-oriented approach should be especially useful for describing knowledge-based skilled performance and vaguely defined tasks, with practical applications to performance measurement, training programs, and intelligent tutors.

General Features

The following features characterize our approach to integrating task analysis methods of personnel psychology and cognitive science:

Model-Based Approach. We employ a general framework of the content of job expertise to guide the task analysis process. This model-based approach provides advantages in efficiency and comprehensiveness. It serves as a guide to the many practical decisions required to adapt the task analysis process to the particulars of a specific job. For example, we use this framework to develop relevant questions to ask when interviewing job experts, to select tasks and contexts for job observation and protocol analyses, and to serve as a stimulus for gathering ratings from job experts.

Representative Sampling. To be useful, applications must be both detailed and comprehensive. To accommodate these different objectives, we employ hierarchical sampling to direct the more resource-intensive, cognitive methods to content areas that are particularly informative about the nature of expertise for a job. This provides a rich account of expertise while making efficient use of time and personnel. As a basis for sampling tasks, we use our model of expertise to provide a framework for collecting ratings from job experts. Comprehensive task analyses of whole jobs help to prevent errors which may result from a narrow focus on limited areas of work, such as examining only the technical content of a job. For many applications, the results of such an approach could be seriously misleading, such as examining only flying skill of commercial pilots while ignoring cockpit communications and management. Hence, the use of sampling techniques and a comprehensive framework of overall job proficiency help to ensure that job expertise will be adequately described.

Cognitive Focus. In contrast to job analysis methods that focus solely on behavior, we explicitly incorporate procedures to identify goals, strategies, pattern recognition, and mental models. Further, tasks should be examined as whole, integrated sequences, so that key mental aspects are not omitted. For example, previous studies of land navigation partitioned this task into procedures for determining location, distance, direction, and so forth. By analyzing isolated skills rather than integrated tasks, the critical decision-making skills of choosing which procedures to use, when to use them, and how to adapt them to the situation were missing from task analyses, training, and evaluation tests. Incorporating the key mental supporting the performance of integrated, whole tasks proved essential for predicting performance. Yet it was given scant attention in existing training, formal job documents, or measures of performance.

Work Performance in Context. From our experience, we find that focusing task analyses more directly on actual performance reveals task and knowledge requirements that are unique and important. For example, we found that performance of technical tasks on the job often interacts with performance of communication, team, and administrative tasks. Additionally, tasks other than primary technical tasks are often de-emphasized or omitted when studied out of the context of the job. For example, information gathered from formal job documents (e.g., training materials, job descriptions), retrospective reports, or laboratory experiments tend to omit communication, team, and organizational-wide tasks and knowledge. In part, these omissions may be due to: difficulties in describing perceptual knowledge, lack of formal descriptions that articulate these requirements, a lack of effective cues that prompt recall of these tasks and knowledge, or to our human inability to describe accurately the contents of our cognitive activities. Whatever the reason for these inadequacies, we find it essential to observe actual job performance to develop complete and detailed descriptions of work expertise.

The Nature of Job Performance

An important challenge for cognitive science methods is to accommodate the complexities of job performance. The work to date focuses primarily on technical knowledge and skills acquired in formal instructional settings. From our perspective, describing the expertise required for proficient performance in work settings introduces an additional order of magnitude in complexity of knowledge content. Job performance involves not only duties other than technical proficiency (e.g., managing work flow, assisting others, communicating effectively), but interactions among these many tasks. In addition to describing the content complexities of job performance, task analysis methods must produce timely, cost effective results to support applications such as intelligent tutors and embedded training.

One strategy for efficiently conducting task analyses and developing applications is to use a well-developed theory to guide the process. We examined two areas of the scientific literature for candidates: personnel psychology and cognitive science. Cognitive science provides rich accounts of the nature of technical expertise. Personnel psychology provides extensive taxonomies of tasks and work proficiencies that can be used to guide job analyses. But neither expertise nor proficiency alone are sufficient to describe job performance.

To accommodate a range of human resource applications, we need to know which tasks get performed and what knowledge supports their effective performance. To achieve this goal, we organized these literatures into a description of job expertise using a task by knowledge matrix, shown in Table 1. This combination of breadth of task dimensions and depth of knowledge structures provides a more comprehensive model of job expertise than can be inferred from either scientific literature taken alone. From the perspective of cognitive science, the model indicates the relevance of a wide range of organizationally important tasks. From the perspective of personnel psychology, the model articulates a rich description of the expertise required for job performance. The integration of task and knowledge taxonomies

from the two disciplines also suggests some relevant issues and new insights about the task and knowledge requirements of jobs, by highlighting: the multi-dimensional structure of task performance; the knowledge required to execute tasks in real, physical environments; and the social/cultural bases of job expertise.

We discuss this model in some detail in this section. Discussing theories about job content may be a departure from descriptions of task analysis methods which focus solely on the data gathering process. However, there are several advantages to having a theory about the nature of job expertise, and to explicitly stating what the theory entails. It suggests relevant issues to scientists (e.g., what is the structure of *job* expertise) and practitioners (e.g., which aspects of performance to emphasize and describe for particular applications). It provides a road map for adapting task analyses to specific jobs (e.g., by suggesting interview probes and sampling strategies). It also helps to standardize certain task analysis procedures (e.g., analyzing and representing performance protocols) by providing an explicit, consistent basis for task analysts' judgments.

The organization of tasks and knowledge depicted in Tables 1 and 2 primarily reflect the mainstream of the personnel psychology and cognitive science literatures, respectively. However, applying this task analysis approach to the computer technician's job and to Marine land navigation suggested to us some departures which we will explain in the text as they arise. Depending on your background and your purpose for employing task analyses, readers may also provide differing organizations of the categories and content within them. We provide brief rationales for our conceptions in the following text.

A Model of Job Expertise. Tasks may be defined as a goal-oriented activity. Human resource practitioners often describe tasks in general form, beginning with a verb. "Determine your present location" is an example from land navigation. The task statement clearly describes the activity, but is general in the sense that it does not tell you *how* the task should be accomplished (by terrain association or by using a map and compass). Nor does it provide a clear performance standard (e.g., within 10 meters), inform you *when* the activity should occur, or indicate *why* certain methods are more effective in particular situations. We use the term "knowledge" to refer to task content addressing how, when, and why tasks are performed.

Table 1
A Task By Knowledge Framework of Job Expertise

	Knowledge Requirements			
Task Categories	**Declarative**	**Procedural**	**Generative**	**Self**
1 Technical tasks (job-specific)				
2 Organization-wide tasks				
3 Teamwork				
4 Communication				
5 Work management				
6 Leadership & supervision				
7 Effort & personal discipline				
8 Skill development				

The framework presented in Table 1 represents a central part of our strategy for implementing cognitive task analyses in a cost effective manner. It informs our hypotheses about expertise, directs our study of tasks, and guides our discussions with job experts. We use it as an efficient, flexible heuristic to focus the task analysis and to ensure that our description of job expertise is complete.

Expertise is highly specific to particular tasks. Fortunately, the contents of many tasks are similar, and the structure of expertise is general across most jobs. For example, within military jobs there are several tasks common to jobs both within and across the military services. These include performing first aid (CPR, dressing wounds, etc.); firing and maintaining weapons; maintaining personal fitness, and military discipline. Other tasks, such as providing supervision and communicating effectively, share a similar structure along with at least some similar content. By structure, we mean that task goals are similar. However, the job importance and specific tactics employed for supervising and communicating may vary across jobs.

In addition to similar task goals, the knowledge required to support those tasks also shares many similarities. For most jobs, knowledge requirements can be characterized in terms of the non-exclusive categories of information shown in the columns of Table 1--declarative knowledge, procedural knowledge, generative knowledge, and self knowledge. Although the detailed content will differ across jobs, the structure of tasks and knowledge for most, if not all, jobs will be encompassed by this framework. Because knowledge content can be classified into different categories depending on its function in a particular task or setting, we do not consider these categories to represent a taxonomy of knowledge. In practical terms, this framework helps constrain task analyses, provides a source for interview probes, and can supply important content (albeit at an abstract level) for elaborating job knowledge.

Task Categories. The rows in Table 1 organize tasks according to similar aptitudes and skill requirements. While there are many ways to organize tasks into meaningful groups (based on relative importance, frequency, co-occurrence, goal similarity, content similarity, etc.), the approach depicted in Table 1 is especially informative to employee selection, training, and performance measurement. These performance dimensions differ with respect to their relative emphasis on cognitive, affective, and motor outcomes[1].

This organization of tasks (i.e., the rows of Table 1) describes the structure of performance across all jobs in terms of eight high level dimensions[2]: technical tasks (i.e., job-specific proficiencies), organization-wide tasks (non-job-specific proficiencies), written and oral communications, teamwork, leadership and supervision, work planning and administration, effort and discipline, and personal skill development. The content within these dimensions are expected to vary considerably across jobs. Further, not all eight dimensions may be required to describe any particular job.

We use this framework to guide task analysis efforts to ensure the comprehensiveness of job coverage. Formal job documents, such as job descriptions, training materials, and so forth frequently omit important duties (e.g., assisting the team, supporting organizational goals outside one's normal duties). Further, these implicit duties often have a large impact on individual and organizational performance

---

[1] This familiar taxonomy is from the training literature (e.g., Gagne, Briggs, & Wager, 1988; Kraiger, Ford, & Salas, 1993).

[2] This taxonomy was adapted from work by Campbell and his associates (Campbell, 1990; Campbell, McCloy, Oppler, & Sager, 1993).

effectiveness. Hence, the task framework provides a benchmark to ensure that all important tasks are explicitly described.

1) Technical Tasks. This group of tasks is comprised of the substantive, job-specific tasks that are central to a job. Designing buildings, troubleshooting computers, tracking and guiding airplanes, and preparing documents are all examples of job-specific technical task content. This performance component typically is the most thoroughly described in job documents. However, as the next section on knowledge components will show, even these descriptions systematically omit certain types of content that are essential to technical task performance.

2) Organization-wide Tasks. In most organizations, individuals perform some tasks that are not specific to their own job. In the military services, these include providing first aid, handling and maintaining weapons, cleaning the area, and so forth. These are duties for which everyone is responsible, in addition to their technical tasks.

3) Team Tasks. Providing support to one's peers and work team is the core of this component. This is one dimension that obviously does not apply to all jobs (e.g., for individuals who work alone). Helping with job problems, providing informal training when needed, and assisting others when they are overloaded are all examples of facilitating team performance.

4) Communication Tasks. Many jobs in the workforce involve making effective presentations, either written or verbal, to other individuals and groups. These communications may be either formal or informal. In addition to message content, proficiency in communicating is a key component of performance effectiveness for these jobs.

5) Work Management Tasks. This dimension includes obtaining and organizing resources; managing time and tasks; and problem-solving and decision-making with respect to resource problems. This dimension does not include providing direct supervision (part of the leadership category) or solving technical problems (part of category 1, technical tasks).

6) Leadership and Supervision Tasks. This dimension involves directing and influencing others, both formally and informally. Modeling appropriate behaviors, setting and motivating others towards goals, monitoring progress, and providing feedback are typical examples of this dimension. This dimension applies to individuals whose work involves groups, whether or not this includes a formal role as a supervisor. Thus, we include in this category effective interpersonal skills such as listening actively, negotiating effectively, resolving conflicts, and so forth.

7) Effort and Personal Discipline Tasks. This dimension reflects the consistency of an individual's day-to-day motivation. It involves the degree of commitment to all tasks, persistence across the range of work conditions (including adverse ones, such as working late, in the cold, etc.), level of intensity, and willingness to expend extra effort when needed. This dimension is distinct from one's technical knowledge, cooperativeness with peers, or communication skills. This dimension also involves stress management skills, the degree of integrity in everyday behavior, adherence to organizational policies and procedures, and standards of personal conduct. It also includes avoidance of counterproductive behaviors such as alcohol and substance abuse, inappropriate absenteeism, theft, and so forth.

8) Skill Development Tasks. Developing skills and knowledge about one's job, organization, industry, and career are essential components of many jobs. This involves acquiring, maintaining, and

evaluating one's own technical, organizational, and personal skills. It includes accepting responsibility for and taking the initiative for training and development, whether the opportunities are formally provided or acquired informally through mentoring, coaching, or self-directed learning.

Knowledge Categories. Knowledge functions in different ways in order to support proficient task performance. We organize this knowledge into four, nonexclusive categories to ensure complete description of content: declarative, procedural, generative, and self. We present a more detailed description of these categories in the discussion that follows, and provide a summary of key points in Table 2.

Declarative Knowledge. With respect to job performance, declarative knowledge involves *knowing what* to do in order to get the job done. This consists of knowing the facts, concepts, principles, and so forth that are acquired and can be remembered (given the appropriate cues), usually in verbal (i.e., 'declarative') form. Additionally, we include in this category two distinctions about declarative knowledge identified by cognitive science research for their relevance to job training and performance: knowledge organization and structure; and mental models.

Knowledge Organization and Structure. Knowledge organization and structure refers to how facts, concepts, and rules get organized in memory. In the early stages of learning skills and job expertise, trainees and novices store the acquired information as a set of loosely related facts. As expertise develops, these knowledge units are grouped for more efficient recall and use. Furthermore, as skills move from a novice to expert level, the basis of knowledge organization changes from surface features (e.g., similar appearance or location) to features based on principles.

Mental Models. Mental models refer to simplified models, or representations, of knowledge that are used in performing a job or communicating to others. An organization of concepts, facts, and rules may serve as a mental model that summarizes large amounts of information about the structure, functions, and interrelationships of an organization, task, or equipment system. A mental model can be as simple as a written outline (e.g., from a training lecture) or it can be visual, such as an organizational chart. They can be employed as heuristics to guide problem-solving and decision-making or as frameworks to help in learning new information. For example, the game of football has been used as a metaphor, or model, of organizational competition. Based on the metaphor, prescriptions such as "play every down" and "when the going gets tough, the tough get going" are generated and applied to the work setting.

Procedural Knowledge. Procedural knowledge consists of *knowing how* to perform tasks. This includes knowing when to use a particular procedure, the steps to perform a procedure, and what standards of precision the task process and product must meet. For many tasks, this may also involve recognizing patterns of cues that signal the next procedure or step to perform. Additionally, this includes knowing alternative strategies for performing the job, and when to apply those strategies to maximize job performance. In sum, procedural knowledge concerns knowing the accepted methods for performing the reasonably well-defined tasks of a job.

Generative Knowledge. In contrast, generative knowledge supports the development of new procedures or adaptation of old ones to new contexts. Hence, this knowledge involves *knowing why* things work--understanding causal relationships, domain principles, and systems knowledge. It differs from declarative knowledge by knowing how to adapt principles and to transfer knowledge from one setting to another. While procedural knowledge consists of knowing how to do a task, generative knowledge involves knowing why the task is done the way it is. Perhaps more to the point, generative knowledge consists of

Table 2
Knowledge Requirements For Performance

Categories of Knowledge	Knowledge Components	Description/Example
Declarative	Semantic & conceptual knowledge	- Facts, concepts & principles
	Knowledge organization & structure	- Content and relationships among concepts
	Mental models	- Streamlined representations of knowledge in visual, semantic, or episodic form
	Concepts	- How conceptual knowledge is organized
	Tasks	- Goal sequences
	People	
	Team	- Special skills of team members, etc.
	Organization	- Organizational structure,
	Boss(es)	- Supervisory goals, work style
	Equipment & Systems	- Enables propagation of action effects
	Environment	- Constraints on choice of methods
	Mission	- Effects on goal priorities
Procedural	Procedure selection	- Selecting optimal procedures
	Goal understanding	- Formulation of goals and their priorities
	Pre-condition recognition	- Identifying whether required constraints are met
	Procedure execution	- Knowing correct sequence of steps
	Goal knowledge	- Knowledge of process precision & outcome standards
	Perceptual knowledge	- Perceiving, recognizing patterns of relevant cues
	Strategic knowledge	- Strategy formulation, selection, & implementation
Generative	Problem representation	- Initial framing & classification of problems
	Problem-solving & transfer knowledge	
	Normative reasoning	- Knowing norms, event frequencies, etc.
	Analogical reasoning	- Reasoning from models in related areas
	Deductive reasoning	- Reasoning from domain principles, rules, etc.
	Inductive reasoning/ Experiential knowledge	- Inferring rules from cases - Acquisition of relational & perceptual knowledge from task practice & job experience
	Systems knowledge	- Enables explanation of status; propagation of effects
	Principles	
	Causal relationships	- Understanding causal relationships in the domain
	Explanations	- Can provide reasons for why events occurred
Self	Meta-cognitive knowledge	
	Control processes	- Scheduling serial tasks; integrating parallel tasks
	Self knowledge	- Possesses accurate perceptions of own skills
	Self-monitoring	- Monitoring own performance processes, outcomes
	Self-explanation	- Generates reasons for phenomena
	Self-directed learning	- Identifying training needs; designing training events; managing learning process

information that supports transfer to different contexts, while procedural knowledge emphasizes application to similar settings.

For example, generative knowledge is brought to bear on defining unstructured problem situations (perhaps the foundation of 'problem representation'). It consists of domain-specific content and processes of knowledge directed to adapting goals and methods to novel situations. To transfer performance to new settings, knowledge is generated by reasoning from job norms (normative reasoning), domain principles (deductive reasoning), well known models in other areas (analogical reasoning), or inferring rules from previous experience (inductive reasoning).

Generative knowledge also includes systems knowledge--the relationships among the parts of a system and how the parts connect to the whole. This knowledge is useful for predicting system status and how effects are propagated among the parts.

Self Knowledge. Self knowledge consists of the meta-knowledge required to plan, implement, and monitor how and when tasks are performed. It also involves knowing what knowledge is needed, how to efficiently acquire it, and how to monitor one's own level of understanding. This includes managing one's own learning process effectively, whether training takes place in formal (i.e., in the classroom or lab) or informal settings (e.g., while being coached or mentored on the job), and whether training is directed by instructors or oneself.

Implications for Task Analyses and Test Design

One intended purpose of the model of job expertise (presented in Tables 1 and 2) is to guide the conduct of task analyses. For example, we should expect descriptions of job expertise to include tasks and knowledge from each cell of the model or an explanation for why it does not apply in this case. In this way, the model provides benchmarks to ensure that task analyses are systematic and comprehensive. As a summary of research and practice on job performance, this model also serves as a reminder that performance is not just 'one thing' (Campbell, 1990; Dunnette, 1963). Performance, and the expertise required to support it, is multi-dimensional. Applications attempting to measure, model, or improve <u>overall</u> performance must recognize the multi-dimensional structure of job expertise. Because portions of job expertise are implicit, care must be given in task analyses to identify it.

The model of job expertise also provides specific guidance for the conduct of each phase of task analysis and test design. For example, the model provides a useful framework for generating interview probes and for classifying performance protocols. It also provides a general framework that can be used to obtain expert judgments for test specifications.

Section 2:
Description of Cognitively-Oriented Task Analysis Methods

Cognitively-oriented task analysis is a collection of procedures flexibly applied to the goal of identifying the task and knowledge requirements of a job. The focus of this approach is to describe expertise associated with job performance. Hence, we emphasize eliciting detailed knowledge that experts actually use while performing tasks, in addition to their (or others') reports about that expertise. The basic approach can be summarized in the five steps shown in Table 3.

Step 1: Plan the Project

Here we comment on two features of project planning especially relevant to our task analysis approach: defining project goals, resources, and constraints; then adapting your methods to meet these considerations.

Project Goals. The goal for conducting task analyses typically involves supporting the development of one or more human resource applications. The nature of the application affects planning by specifying the scope and depth of information that needs to be obtained. For example, developing performance measures requires comprehensive coverage of a job at a moderate level of detail. In contrast, developing intelligent tutors requires fine-grained details, but often is restricted to technical knowledge.

Table 3
Cognitively-Oriented Task Analysis

Activities	Steps
1. Plan the project A. Identify application goals, resources and constraints B. Define approach	• Interview senior management • Design sampling plan • Collaborate with a job expert • Select methods
2. Analyze tasks	• Interview job experts • Review job & training documents • Use task x knowledge framework • Gather performance examples • Develop task questionnaire
3. Identify diagnostic tasks	• Obtain expert ratings
4. Elicit detailed job knowledge	• Conduct protocol analyses
5. Represent job expertise	• Develop plan-goal graph • Develop task by knowledge matrix

In addition to specifying the application, you also need to identify how the application will be used. For example, job knowledge tests can be used to diagnose individual performance, predict proficiency, promote the best qualified candidates, or to evaluate the effectiveness of training programs (vs. assessing the student). Each of these uses affects how the information is gathered and how it will be used to develop an application. For example, which tasks get selected for more detailed study will differ between uses involving predicting job performance and evaluating training programs. Greater emphasis will be given to tasks showing high performance variability for the former use, and more emphasis will be given to organizational importance for the latter use.

For example, in the computer technician's job, loading tapes to record ship operations data is organizationally important, but is a task which shows very little variability in performance across technicians.

Because this task is central to performance and is formally taught, tests designed to evaluate training should include assessment of this task. However, if the objective is to predict performance, questions assessing tasks with little or no performance variation will add little to your knowledge about differences among technicians' performance. Instead, assessing technicians' capability to train themselves will probably be more useful because there is substantial variation in performance of this task.

Inevitably, specification of application goals and uses will involve discussions about what aspects of job performance are relevant. For purposes of task analysis planning, these discussions should focus on three topics: people, tasks, and contexts. The number and range of possibilities for these three factors need to be specified to ensure that task analysis results will reliably generalize to your application goals.

Using our land navigation task as an example, it was important to conduct task analyses in at least two different environments (i.e., contexts) of mountains and forested plains. As a result, we identified important differences in strategies, methods, and expertise across these environments. In other military settings, specifying the range of relevant war and peacetime scenarios involved in job performance will be similarly important to effective planning.

The primary implication for planning task analyses is to determine an adequate sampling plan across the three factors of people, tasks, and contexts. For example, with respect to people, we found several stable differences in nominal job experts. These included differences defined by strategy preferences and by recency of experience. That is, we defined and studied a group of individuals who were nominated as experts owing to their previous experience, but whose current skills had deteriorated. Including this group of 'decayed experts' in our task analyses provided us with additional insight into the nature of expertise for this task. At minimum, sampling across the most salient distinguishing factor(s) in each class of people, tasks, and contexts allows you to estimate the range of expertise associated with job performance. Some relevant factors will be discussed in the next section on task analysis.

Step 2: Analyze Tasks

The goal of this phase of task analysis is to develop a complete list of the duties and tasks involved in a job. We employ interviews to achieve this goal, supplemented by a structured approach to gathering examples of job performance (i.e., the critical incident method; Flanagan, 1954). While not a required step in our approach, it is an especially useful method for extending the task analysis to tasks and contexts that may not be available to job observation (e.g., due to safety or cost constraints). The outcome of these methods will be a questionnaire that can be used to target additional task analysis efforts for describing job expertise. We begin this section by extending our model of job expertise, then showing how it can be used to assist the task analysis process.

Using the Model of Job Expertise. The model provides us with some initial hypotheses about the content of expertise. In applying the model to task analyses, we comment on three aspects of tasks that may affect the nature of job expertise: task content, task characteristics, and job context.

Task Content. When job experts provide retrospective reports about performance, they frequently have difficulty recalling and reporting all of the tasks that they perform. They tend to omit tasks that are not part of the technical content of their job or are not included in official job documents such as job descriptions or training manuals. Unfortunately, these omissions too often represent significant portions of the job. However, the framework suggests useful probes and cues to assist job experts in describing their work.

Using a computer technician's job as an example, it was common for job incumbents and supervisors to discuss their job in terms of operating, maintaining and repairing computers (i.e., technical task proficiency). With some additional probing, they were able to describe a wide range of additional activities that they performed, including participation in collateral duties (e.g, tasks related to physical plant maintenance, safety, and security), training and assisting team members, communicating information throughout the organization, and planning and administering their work (organizing maintenance schedules, ordering parts, etc.).

Although formal training is not provided for such activities, proficiency in some of these tasks appears strongly related to supervisory assessments of overall job performance. Further, performance on these tasks often interacts with performance on technical tasks. Thus, capturing this information is important to the development of job aids and performance measures that are intended to support or assess overall performance.

Table 4
Effects of Task Characteristics on Knowledge Requirements

Task Characteristic	Knowledge Requirements Affected
Importance	Goal knowledge & organization; task strategies; procedure selection
Time, outcome pressure (maximum vs. typical)	Goal knowledge & organization; task strategies; procedure selection
Goal focus (speed vs. accuracy)	Goal knowledge & organization; task strategies; procedure selection
Goal difficulty, complexity	Declarative knowledge; system knowledge; pattern recognition & procedure selection
Task consistency	Proceduralization of knowledge function vs. pattern recognition & procedure selection

Task Characteristics. In addition to content, there are other task characteristics that can affect the knowledge requirements of a job. In Table 4, we identify several of these and briefly characterize their impact on job knowledge. In fact, characteristics such as importance, difficulty, pressure, and consistency can affect both the content and processes by which individuals perform their work.

The amount of pressure on task performance varies across tasks and situations. The repair of ship-board computers when technicians are in port requires knowledge of diagnostic procedures and a moderate level of motivation. Repairing the same problem when under enemy fire not only requires increased speed and attention, but knowledge of how to optimize high priority tasks and satisfice low priority tasks.

Each of the task characteristics presented in Table 4 represent sources of potentially revealing information about the nature of expertise for a job. We evaluate their potential first by asking questions related to these task characteristics in initial interviews, then later explore their relevance through job

observations. Additionally, understanding the relative organizational importance and amount of performance variability in each class of tasks may provide you some important clues for productively focusing the task analyses (e.g., using protocol analyses) and for improving existing applications.

Task Context. Contextual factors often exert their influence through the changes they impose on task characteristics. The previous example concerning navy computer technicians illustrates this point. The level of security threat, routine steaming or in battle, impacts task pressure and goals. Contextual factors such as the environment (e.g., in port vs. at sea) and organizational mission can impact knowledge requirements in similar ways. Other contextual factors, such as the nature and amount of resources available, may have their impact through the job performer's selection of goals and the procedures used to satisfy those goals.

The model of expertise displayed previously in Tables 1 and 2 is intended to provide a good starting point for identifying the nature of expertise in a job. In this section, we articulated it further by adding considerations of task characteristics and task context. The categories and content of this model of expertise are general, domain independent, and abstract. However, job expertise is domain specific. Hence, the model is intended to provide direction for elaborating the details of job expertise, and to guide adaptation of task analysis methods to your particular situation. We illustrate this use of the model in the following descriptions of our task analysis methods.

Interview Job Experts. The primary goal for initial interviews with job experts is to define job duties and tasks. Additionally, we use this occasion to identify potential differences in expertise, tasks, and contexts that should be incorporated into the sampling plan for more extensive knowledge elicitation efforts. Finally, we also use these initial interviews to introduce the project to job holders, answer their questions, and encourage their participation. We find that time and interest invested early with these job experts yields essential ongoing support and cooperation during the project. Be aware that your goals may be considered mere overhead for your job experts. Take the time to explain how your project will benefit them and their work.

Interviewing three to five job experts is generally sufficient to arrive at a converging set of major job duties. Experienced job incumbents (e.g., with 3 or more years experience), or supervisors who have extensive experience performing the job, are appropriate as job experts. Where possible, we select interviewees who are both competent performers and verbally fluent.

Table 5
Organization of a Job Analysis Interview

1	Project introduction
2	Background information
3	Open-ended questions about job
4	Follow-up probes
5	Informal ratings of task characteristics
6	Summary
7	Close

One organizational scheme for the interview is shown in Table 5. These interviews are semi-structured and take about one, to one and a half hours, with each interviewee. We usually begin by describing

the purpose of the project and the importance of their contributions. The primary focus of the interview is on developing a general, yet complete list of all activities comprising the job. Hence, the use of open-ended questions is recommended. For example, the following questions may be useful.

"What do you do on a 'typical' day?"

"What are the major goals and activities in your work?"

Table 6
A Guide for Interview Probes

Topic	Example Probe
Performance Categories	
Technical proficiency	Please describe your primary job duties.
Organizational-wide proficiency	Outside your primary duties, are there other tasks you perform?
Teamwork	What roles, if any, do you perform in work teams?
Communications	What types of written and verbal communications do you do in your job?
Work planning & administration	How do you plan and administer your work?
Leadership & supervision	In what ways does your work require you to influence or guide others?
Effort & personal discipline	In what ways does your work require you to persevere, work late, or expend extra effort?
Training & development	Please describe areas for which you train or update your skills.
Task Characteristics	
Importance (to organizational goals)	Please rate the relative importance of the duties we have just discussed.
Pressure (maximum vs. typical)	Which duties/tasks are performed under pressure of time or outcomes?
Goal focus (speed vs. accuracy)	Is speed or accuracy primarily emphasized for this duty?
Complexity	Which of these duties/tasks are more difficult, requiring extra thought before responding?
Consistency	Which tasks can be performed in a relatively routine way?
Task by Person Considerations	
Performance variability	Which duties/tasks produce the most variability in performance?
Time spent	How much time do you typically spend on each of these duties/tasks?
Contextual Factors	
Organizational goals/mission	What are the organizational goals or missions that are especially relevant to your job?
Work group collaboration	For which duties/tasks do you depend on others for assistance?
Equipment	What equipiment do you use to accomplish your job?
Resources (mentors, job aids)	What other resources assist you in your work?

The use of open-ended questions and unobtrusive follow-up probes is recommended because capturing the interviewees' terminology and organization of tasks can provide insight into their conception of job performance. We present some examples of follow-up probes in Table 6. It should go without saying that taking careful notes and/or recording these interviews is essential. You won't remember as much detail as you think you will.

In addition to clarifying and expanding descriptions of job activities, follow-up probes are usually necessary to assist the interviewee in recalling and articulating job activities. Job experts' conceptions (and

verbalizations) about their job are frequently dominated by the representations found in formal job descriptions, performance appraisal forms, and training materials. Unfortunately, it is often the case that these formal descriptions are substantially deficient. These documents tend to describe only technical task performance while omitting such organizationally important activities as providing team support, communicating with organizational members, providing informal training or supervision, and so forth.

After developing a thorough picture of job tasks, we probe for information about the effects of task characteristics and task context. This information can also be gathered by asking the interviewee to rate each of these characteristics.

Using interview notes, we consolidate the information into a representation of task content, structure, and contexts. This often takes two forms, a task list and a graphical representation of task structure (e.g., the plan-goal graph discussed in a following section).

Incorporate Information From Job Documents. For most jobs, there exist a variety of sources that can be used to further delineate the tasks and duties outlined in the initial interviews. These materials include training manuals (e.g., instructor guides, training path charts, PPP tables), technical reference manuals, job aids, performance appraisal forms (e.g., Personnel Qualification Standards), job descriptions, and mission statements. The goal of this activity is to refine the list of tasks and activities that comprise the job. Any noticeable differences between representations of the job found in job documents and from interviews is a potential source of content for differentiating among levels of expertise.

Gather Performance Examples. Another way to develop a detailed description of the job is to collect performance vignettes from job incumbents and supervisors. This supplement to the other methods is valuable for several reasons.

First, it often identifies knowledge that is important to performance, but that is not typically described in job documents or readily articulated in interviews. By focusing directly on performance, it provides improved access to knowledge developed from job experience. Identifying this 'implicit' knowledge appears important to adequate characterizations of expertise.

Second, it extends the task analysis by incorporating performance incidents from a wide range of situations and contexts. We employed this method to gather information about performance in environments that were not practical to observe directly (e.g., land navigation in desert and tropical areas; electronic repair during combat conditions).

Third, examples of actual performance provide a rich source of information about the performance context (goal interactions, resources used, constraints encountered, errors committed, etc.). In addition to insight into complex performance, these vignettes provide the basis for scenarios that can be incorporated into applications such as training and performance measurement. Finally, the application of this methodology potentially involves most job incumbents and supervisors. Their participation in the early phase of task analysis provides the opportunity to increase their understanding and support for the application to be developed.

Description of the Critical Incident Method. The methodology is an adaptation of the critical incident method (Flanagan, 1954; Smith & Kendall, 1963). The method involves providing job incumbents and supervisors with a structured approach to writing about examples of performance that they have directly observed (their own or others). An example of a completed form is provided in Figure 1.

The key to writing effective performance examples is to provide systematic training for the individuals who will write examples. Training consists primarily of providing examples and opportunities for practice with group and individual feedback. The tendency is for participants to provide abstractions, summaries, or prototypes of performance rather than specific, actual events. The power of this method rests on its specificity. Thus, training is essential to ensure that participants understand the level of detail required, and the format and purpose of the exercise. Training takes about 30 minutes.

Depending on job complexity and the nature of the application to be developed, 100 to 600 incidents may be needed to adequately characterize the job (e.g., to cover the range of performance from novice to expert for 6 to 10 different dimensions of performance). Participants produce about 3-5 incidents per hour and can remain productive for about 2 hours. Hence, 20 individuals in a three hour group session (including training) could produce about 150 to 200 performance examples. Individuals who are verbally more fluent and who possess more job experience tend to write more, and better, incidents.

Two hour sessions are not uncommon, given practical constraints on access to personnel. Sometimes, only short intervals are available. For these situations, the task analyst should verbally interview the job expert, using the critical incident format. This approach has been reported to be effective for knowledge engineering purposes (Klein, Calderwood, & MacGregor, 1989).

PERFORMANCE EXAMPLE FORM

1. What were the circumstances leading up to the incident?
   Data recording for CEC missile shoot. The ACTS RD-358A was showing a multiple dead track error and wouldn't dupe a tape.

2. What did the individual do that made you believe he was a good, average, or poor performer?
   After troubleshooting and cleaning the tape drive heads, the technician observed that the file reel was not gripping the tape properly. When the tape moved forward, it slipped causing a multiple dead track error. The tech then replaced the file reel hub with a new one.

3. What was the outcome, or results of this incident?
   We were able to reduce and duplicate tapes during the missile shoot.

4. Circle the number that best reflects the correct effectiveness level for this example.

0	1	2	3	4	5	6	7	8	9	10
ineffective			less effective		about average		effective		extremely effective	

5. This performance incident is relevant to what performance category(ies)?:

   Repair equipment

6. This incident is descriptive of what job? Computer Technician

Figure 1. A completed performance example form for the computer technician job.

The follow-up questions for each incident minimally should describe the pre-conditions (events leading up to incident, resources and constraints, critical cues, etc.), actions taken, and outcomes. Depending on the task analysis purpose, other probes may prove useful. Queries about specific task goals, other options available, decision criteria, and how changes in situational factors would have affected the actions or outcomes can enrich performance examples.

Conceivably, many other probes could augment the information gathered. However, avoid overwhelming the participant with queries. The effectiveness of this method depends on having participants recall specific incidents that they observed. While people appear capable of reliably recalling circumstances, actions, and results that unfolded over many seconds, minutes, or longer, we caution that their reports on their own (or others) cognitive processes (thoughts, strategies, cues perceived, etc.) are unreliable (Ericsson & Simon, 1984; Nisbett & Wilson, 1977). If such information is gathered, it should be considered only for generating, not for confirming, hypotheses about the nature of expertise.

Analysis of Critical Incidents. The first step in analyzing the performance examples is to organize the incidents into categories based on similarity of content. The typical basis for judging similarity is task content (e.g., problem-solving, communications, safety, operating equipment, etc.), although other bases may also be appropriate (e.g., goals). This sorting of incidents into categories is usually carried out by the task analysts. It provides another source of useful insight into the job and the expertise required for performance. When all the incidents have been sorted, then category names and definitions are developed based on the content of the performance examples in each category. This often results in some re-sorting of incidents into other categories. Also, it is common practice to edit complex incidents into several, more simple and homogenous incidents.

As a check on the reliability and meaningfulness of the resulting organizational scheme, the next step involves having several job experts sort each incident into one of the categories based on the category labels and definitions. From this data, indices of agreement for each incident can be computed. Incidents with low agreement are then either deleted or edited to fit the most appropriate category. Inter-rater reliability between the job experts can be computed as one indication of the meaningfulness of the categories.

Once the incidents and categories have been established, then have job experts rank order the incidents within each category according to the level of performance effectiveness displayed. This can be accomplished by having each expert provide an absolute rating of effectiveness for each incident.

The scaled incidents are useful in several ways. They inform you of the range and variation of performance within each performance dimension. Also, they provide another source of information about the tasks and expertise comprising job performance. This description of performance should be compared to the task list prepared in previous steps of the task analysis to see if any new tasks or expertise should be added.

In sum, gathering performance examples provides a unique source of information about job performance. Unlike job documents and employee interviews, this method focuses job experts on specific, detailed accounts of critical performance incidents. Distinct from protocol analyses, it provides accounts of performance occurring in circumstances that might not be available to observation due to safety or cost constraints.

Step 3: Identify Diagnostic Tasks

Tasks that are more informative, or diagnostic, of expertise are targeted for further analyses. Because detailed task analyses are time consuming to conduct, focus these efforts on the tasks where expertise makes the most difference. To accomplish this objective, we obtain ratings from job experts on two tasks and then use this information to develop a sampling plan to guide our knowledge elicitation efforts.

Rating Tasks and Knowledge. First, we ask them to estimate the relative diagnosticity of task and knowledge categories for the job. Second, we have them judge the diagnosticity of tasks within each task category. We accomplish this by having them rate the relative importance and performance variability of each task. Taken together, information from these two rating tasks provides a clear rationale for targeting our knowledge elicitation efforts.

Selecting and Training Raters. To ensure the quality of the ratings, we specify three knowledge requirements for those selected as raters: (1) technical expertise in the subject area of the ratings; (2) extensive experience in observing performance under the range of conditions and contexts for which the ratings will be made (i.e., knowledge of performance norms); and (3) thorough understanding of the rating task. Where possible, we attempt to obtain the participation of 5 to 10 experts for these ratings tasks.

Description of Rating Tasks. We typically conduct both rating tasks in a single session of about 90 minutes. We begin the session by describing the project purpose and communicating its importance to the job experts. This helps to ensure their interest and commitment to providing useful information.

Category Ratings. An example of the category rating task is shown in Table 7. The table presents a matrix of categories of job duties and knowledge for the computer technician job. The rating task consists first of having job experts assign percentages, summing to 100, to each row of task categories to reflect the extent that performance on these tasks exhibits job expertise. When our application involved developing a job knowledge test, we also stated this another way. The experts were asked how they would weight test content to give them optimum information about overall job proficiency. The assigned weights should then reflect how informative performance in each task category is to overall job proficiency.

Table 7
Description of Expertise for Computer Technicians

	Knowledge Categories					
Job Duties	Principles & Concepts	Procedure Selection	Procedure Execution	Goal Knowledge	Pattern Recognition	Percent Diagnosticity
1 Data recording & reduction						14%
2 Monitor & maintain equipment						20%
3 Repair equipment						24%
4 Clean equipment, workspace						4%
5 Assist work team						7%
6 Communications						7%
7 Work planning & administration						6%
8 Ship-wide duties						2%
9 Maintain personal effort & fitness						7%
10 Training oneself						10%
Percent Diagnosticity	15%	21%	19%	20%	25%	100%

Averaged over all raters, assignments of higher percentage indicate that the task category is relatively more important and has greater performance variability (i.e., requires more expertise) than the other task categories. If there is little performance variability in a task category, or the category is relatively unimportant, then it should receive a low rating because it will provide comparatively less information about overall job proficiency.

Similar ratings are then made for the categories of knowledge in each column. Ratings on these categories indicate the job experts' view about how each type of information content impacts performance in their job. In essence, the job experts estimate the relative importance and amount of information for each type of content. Each of the knowledge categories reflect types of information that have been shown to be generally important to job expertise. We take special care to describe, illustrate, and discuss the definitions for each category of knowledge with the job experts. We accomplish this by briefly defining the category, providing examples from their job, then discussing each category with them. It is important to ensure that

they thoroughly understand the rating task before proceeding because this way of conceptualizing expertise in their field will probably be new to them.

After independent ratings are made by each job expert, then we ask each expert to present their ratings to the group along with a brief rationale. After all have presented and the results tallied on the board, we discuss any discrepancies that occur. Following the discussion, we have the experts make the ratings again. We collect both sets of ratings, but use the last set of ratings for our analyses.

Task Ratings. The second set of ratings provide information about the tasks that most clearly display job expertise. In this exercise, job experts are asked to rate two characteristics of each task: 1) its importance to organizational effectiveness; and 2) the extent of performance variability observed for the task. These ratings are made independently by each expert on forms we provide. After averaging across raters, we multiply the two ratings for each task to obtain an index of the relative diagnosticity of tasks. We use the resulting information to prioritize our implementation of knowledge elicitation, the next phase of the task analysis.

Reliability of Expert Ratings. In our experience, job experts have reported that these ratings are meaningful and straightforward to make. The correspondence among their ratings supports their statements. Inter-rater reliabilities are moderately high--.86 for the category ratings and .78 for the task ratings.

Developing a Sampling Plan. The results of these rating tasks provide a quick snapshot of experts' views of the expertise required for the job. This serves two purposes. It targets our efforts in the next step of task analysis--eliciting job knowledge. It also provides a framework for the development of applications, such as providing specifications for job knowledge tests, or priorities for curriculum revisions. This use of task analysis results will be illustrated with an application to test development in Section 3.

For most applications, you will need to ensure that the description of expertise you develop is reasonably complete and accurate. You will also need to balance this objective with the costs in time and resources of achieving it. The solution to this dilemma is to gather protocols from a well-chosen sample of the people, tasks, and contexts that comprise the job.

You will soon discover that experts differ in their expertise, their approach to the work, and in their definitions of who is an expert. Fortunately, these differences tend to cluster systematically into groups. Observing a variety of job incumbents, when available, provides valuable information about variations in task strategies and methods. In addition to observing people at a variety of proficiency levels (e.g., experts, journeymen, and novices), observing individual differences within proficiency levels also provides insight into the nature of expertise for the job. For example, sometimes differences exist between experts who have served as instructors versus those who haven't. Consistent differences may also occur in work strategies. In our work in land navigation we found two consistent styles of navigating--by using terrain association and by map and compass. After defining categories of expertise, then you can select individuals from each group to serve as subject matter experts. As a final note, you may also find it useful to actually test their level of expertise. Referral by others is an expedient but not always reliable criterion of expertise.

For sampling tasks, we propose that you employ a hierarchical sampling plan using task diagnosticity ratings to prioritize task selection. This sampling should include opportunities to gather information from each of the major task categories that comprise the job. Care should be taken when defining and sampling tasks to include all essential elements of the task. As mentioned previously, tasks

should be studied as whole, integrated sequences in their natural context to ensure that all essential elements are identified.

Ideally, you can gather performance protocols across a sampling of the major contexts, or environments, in which performance occurs. For example, in the land navigation task, we gathered protocols in forested, flat terrain and in mountainous terrains. The differences in expertise and performance across these two environments were considerable, and well worth the additional resources required to study them. In addition to representatively sampling environments, you will also want to consider other types of contextual differences. We mentioned some important task characteristics earlier in this report (e.g., consistent vs. inconsistent tasks, maximum vs. typical demand) that may deserve attention in selecting contexts for task observation. In military settings, this certainly requires attention to different levels of combat alert, types of threat, and so forth.

Gathering performance protocols across a representative sample of people, tasks, and situations will rarely be completely possible. One strategy for addressing deficiencies in your sampling plan is to gather performance examples, as described previously in this report.

Step 4: Elicit Detailed Job Knowledge

The purpose of knowledge elicitation is to identify the information job incumbents actually use for performing their job. In some ways, this is a straightforward task. For example, it is fair to assume that your physician must possess knowledge of anatomy, biology, pharmacology, and so forth. You could add to your list of knowledge requirements by examining standard texts used for training physicians.

However, what makes knowledge elicitation a much more intriguing and challenging endeavor than simple list making is that so much of what contributes to medical expertise has been learned from experience. As in other jobs, physicians acquire their knowledge from a variety of sources--their own experience in internships and residencies, talking with colleagues, mimicking expert performance, reading journals, and by reflecting on their knowledge and experience. Consequently, much of what is important about their knowledge is implicit. Asking them direct questions will not provide you with a satisfying account of their expertise. To draw out this implicit knowledge, you need to expose the expert to tasks that require this knowledge to be used and made explicit.

The primary methods we use for knowledge elicitation involve obtaining and recording the verbalizations of job experts (and novices) during performance of actual job tasks in their natural context. Descriptions of expertise using these verbalizations as data indicate the knowledge requirements of the job. By examining the contents of current awareness, we gain insight into what information is actually used to perform their job.

The assumptions underlying these methods are that: (1) people can reliably report the content of their current awareness; and (2) verbal reports consist of the information that is actually used for task performance. Based on considerable research, we also assume that people's explanations of their performance and their reports about past experience are often inaccurate. Hence, the emphasis in these methods is to have job incumbents (we'll call them subject matter experts, or SME's) 'think aloud' while performing a task, rather than explain what they are doing after the fact.

Gathering Performance Protocols. We employ three related methods for knowledge elicitation: protocol analyses, coaching, and analyses of team communications. All three methods involve having you observe and record the verbalizations of your subject matter experts (SMEs) as a way of learning about the content of the

mental activity required to perform the job. All of the methods require you to interpret the observations after they are obtained. The methods differ primarily in the degree of influence that you or other participants exert on the exchange.

- **Protocol Analyses**. Protocol analyses involve obtaining verbalizations from SMEs while they work alone (Ericsson & Simon, 1984). Using this method, the person is asked to "think-aloud", thereby providing verbal markers of the contents of working memory. The role of the task analyst is only to prompt the SME to continue verbalizing.

- **Coaching**. Using coaching (Gelman & Gallistel, 1978), the SMEs provide you with instructions for performing a task while you execute it in their presence. Unlike your usual role as a good listener, you are not trying to fill in lapses in completeness or guess the intentions of the SME. Your role here is to encourage SMEs to articulate their instructions thoroughly.

- **Team Communications**. Ordinary communication within a team also provides a verbal record of cognition (Orasanu & Fischer, 1992). Your role here is diminished because the team members prompt each other to communicate. But the team members' awareness that they are being observed may still influence their behavior.

General Description. For all three methods, the purpose of your interaction is to keep your subject matter experts talking, using their typical task language. We find it essential to ensure that the SME feels comfortable about making, indicating and repairing mistakes. Everyone makes mistakes. In fact, mistakes are typically more informative about cognition than correct performance. Further, the ability to detect and repair mistakes is an essential component of expertise.

Subject matter experts (SMEs) are nearly always eager to assist you and to impress you with their knowledge. When you elicit job information from SMEs, the demeanor you exhibit influences their responses. Though you cannot eliminate this influence, you can attempt to reduce its negative consequences. A serious negative consequence is that your SMEs will edit their accounts, providing a view of the task domain that they believe meets your approval. An edited account of the job will interfere with your objectives of accurately describing job expertise.

A judgmental demeanor that emphasizes status differences between you and the SME, or a refusal to converse with the SME under the guise of preserving objectivity, will probably reduce the amount that you learn from the job expert. For similar reasons, avoid interactions that require the SME to report on their domain in the foreign language of your theory of task analysis and cognition. For example, do not ask SMEs to categorize their comments as either declarative and procedural knowledge.

Hence, it is important to consistently communicate respect for, and interest in, what your SMEs may be saying. Even if your interest is not genuine, you can still interact as if it were genuine. Perhaps your interest will be genuine in the next topic your SME raises. Another approach to handling the effects of your influence is to reduce the importance of your approval. For example, acknowledge that you and the informant are both experts, but in different domains. You are an expert in task analysis. The SME is an expert in the domain you are analyzing. A novice SME is likely more expert in the domain than you are. And even if this is not accurate, you can still interact as if it were accurate.

In summary, all of these methods assume that the task analyst is as passive as possible within the limits of friendly interaction. The task analyst primarily intervenes only to prompt verbalization (job experts often forget to express their thoughts), but not to suggest interpretations of the information.

The Use of Scenarios. While there are advantages to gathering protocols of actual work performance, this is often not practical. In addition to cost and safety constraints, this practice could result in observing a very limited and unrepresentative set of task performances. Consequently, we typically gather protocols of task performance under a simulated set of conditions. Typically, we construct a set of scenarios that incorporate the tasks and contexts that best display job expertise. These scenarios consist of a few paragraphs that describe important features of work situations. To develop scenarios, we use information from critical incidents gathered in step 2, the diagnostic priorities established in step 3, and assistance from our collaborator SME.

For example, while studying land navigation we constructed scenarios that described the mission (e.g., deliver supplies to an infantry patrol within the next hour), context (in hostile territory), environment (mountainous terrain), and situation (you are the unit leader and must plan the navigational route). After first describing project goals and instructions for the data gathering session, we provided SMEs with a scenario, then had them begin thinking aloud while they performed the task. Although we used simulated scenarios, we observed and collected protocols of performance in its natural context. For land navigation, this involved navigating in large wilderness areas.

Alternative Methods of Data Gathering. For practical reasons, we employ other methods to capture this information when it is not feasible to do so using protocol analyses. For example, following task completion some retrospective probes can be employed to further clarify the job knowledge used. Queries about goals, perceptual cues and patterns, decision options and criteria, performance standards, and so forth may prove useful in extending your understanding and modeling of job knowledge. At the end of a session is also a good time to request clarification, if you sense that you do not understand the meaning of an SME's account. We employ these procedures at the end of the session to avoid biasing the SME's account.

To probe for implicit goals, we also might ask SMEs what they would do under hypothetical situations. Another approach is to conduct more in-depth interviews about expertise used in past situations. A variant of the performance example method discussed earlier, this approach has been shown to be an effective knowledge elicitation strategy (Klein, Calderwood, & MacGregor, 1989).

Although retrospective reports are limited by inaccuracies of memory and faulty inferences, the advantages of their use can exceed the risks in some situations. We can extend our understanding by gathering information about the expertise involved in contexts and tasks other than those from which we gather protocols. These self-reports can provide a rich source of information and ideas about job expertise. As with protocol data, hypotheses about job expertise based on these data are tested through evaluation of the application that is developed.

Documenting Performance Protocols. We recommend videotaping all knowledge elicitation sessions. The videotape captures visual aspects of the task as well as the experts' verbalizations. You will likely require this record of the task setting in order to interpret the verbalizations (particularly pronouns). The record may contain pointing and examples of task-related physical actions that are not indicated in the verbalizations. You may also add markers to the visual or auditory record to assist your later interpretation. For example, when we videotaped electronics repair activities we called out and tagged the page numbers of documentation

as they were accessed. The use of portable video recorders and lapel microphones will improve the quality of your recording and the ease of understanding subjects' verbalizations (e.g., over the din of extraneous noise).

Analyzing Performance Protocols. The purpose of analyzing performance protocols is to develop hypotheses about job expertise. Protocol analysis is an ongoing process of modifying your hypotheses as you encompass more observations within your theory of domain expertise. In our approach, protocols are not analyzed to test hypotheses. Hypotheses can be tested later by evaluating the application that you develop.

Protocol analysis begins with a preliminary decomposition of the domain into goals and sub-goals. With this initial structure, you can then identify individual methods and apply them to the goals in effect. The purpose of this aspect of the analysis is to identify the goals and methods by name. Also, your analysis of the protocols ought to indicate interactions among methods and goals, or the side-effects of one method or goal on the feasibility of another method or goal.

Your first decomposition won't be adequate; your tenth decomposition won't be perfect either. But over time, new observations will require increasingly minor modifications to your representation of expertise, ultimately merely comprising the addition of a new method for achieving some goal you had already represented.

The strategy for analyzing the performance protocols involves three activities, often conducted in parallel with each other, and with the activity of developing representations of job expertise (step 5). These activities are: (1) preparing the protocols; (2) identifying and inferring the goals of the work activities expressed in the protocols; and (3) determining the methods, or plans, used to address the goals. We next present some background and explanation for these activities. This process is illustrated with an example from an electronics repair task in Appendix A.

Protocol Preparation. Depending on the amount of time and resources available, the process of protocol preparation can range from formal and detailed to very informal analyses. Each step in the analysis of protocols can be enormously time consuming. For example you will need to organize and prepare the videotaped data.

In an informal review, you may decide to simply take notes on the observations or construct a knowledge representation directly from watching the videotapes. Alternatively, you can transcribe a portion of the protocols more thoroughly and use the remaining videotapes to refine your preliminary task analysis.

Typically in our approach, we transcribe the verbalizations and add some descriptions of the actions we observe. This requires about eight hours of transcription for each hour of tape. The protocols we transcribed for electronics repair involved this level of activity. In addition, we collated the protocols with the technical documentation that SMEs used.

At the formal extreme, someone who is interested in communication might spend weeks or even months on the same hour of tape, encoding every nuance of the verbalization, including the emphases on words, the pauses in speaking, the processes by which other participants interrupt or encourage the speaker, etc. In other words, creating the representation of the observations is an analysis in itself. It reflects the study purpose and theoretical predispositions about which behaviors are significant.

Goal Identification. Goals identify the purposes of action. Identifying the goals of the work domain is an essential part of the analysis process and must not be compromised with unmanageable time

constraints. One reason that establishing the goals of work merits special attention in task analysis is that they are often implicit in behavior and instructional documentation. Thus, these goals must be inferred from the data and made explicit.

Consider the task of cooking. The purposes of cooking are actually rather complex. We cook to alleviate immediate hunger and recover energy. We also cook to destroy bacteria, facilitate digestion, and enhance health. Finally some of us cook to serve or entertain others, or to create an unusual taste or appearance. Purposes also include a great deal of social and cultural "common sense". A cook who creates a visually appealing, tasty, nutritious meal in a timely fashion is less than successful if the kitchen burns down in the process. Although you may never observe a cook generate this sorry outcome, there's no doubt that the cook takes precautions to guard against fire in every cooking episode. Thus, multiple goals influence the methods we use to cook.

Some of these goals will be evident in protocols of typical performance. To reveal other goals, it is often necessary to modify the constraints of the task you are observing. This frequently involves constructing scenarios that can be provided to SMEs as instructions at the beginning of protocol gathering sessions. For example, in our land navigation study, SMEs never got lost. To examine their performance under these conditions, we had to impose this situation in a scenario. As we develop ideas about the various goals operating on performance, we vary the scenarios to expose whether these goals exist.

Whether protocol documentation is done formally or informally, analyzing protocols requires the most training and experience on behalf of the task analyst. Primarily, this expertise consists of an in-depth understanding of cognition and its contents. This knowledge supports the task of classifying protocol content into goals and methods. The model of job expertise presented in Table 1 represents an initial framework suitable for this task.

Determining Methods. Methods identify the various procedures used for achieving goals. Methods are typically more evident than goals in the protocol data. We identify and label as a method, statements about actions taken. In fact, most of the protocols involve methods. Protocols are segmented into different methods when either of two conditions are met: (1) the protocol segments express plans for accomplishing different goals; or (2) the protocols describe alternative plans for accomplishing the same goal. An additional set of cues alerting you that distinct methods are involved is when a method, or plan, involves using different tools or different features of the task environment. A goal must be inferred whenever two or more methods are identified for achieving the same purpose. We proceed through the protocol data using these rules until we are confident that all statements can be reliably classified into one of the existing goals or methods that we have named.

Step 5: Represent Job Expertise

There are several approaches available for organizing and representing the information gathered throughout the task analysis process. We describe two of them, each of which has certain advantages. The task list adapts easily into a questionnaire format for gathering additional data from job experts. The plan-goal graph method provides a graphical depiction of relationships among tasks. This provides a basis for inferring job knowledge related to task selection and task interactions. These methods possess complementary advantages, so we use both.

Task Lists. This format involves developing a list of tasks and knowledge, organized at four levels of abstraction. This format is straightforward and easy to use. Information from various sources can be integrated and recorded in this form using a typed or database format. This organization of task information

lends itself readily to incorporation into a questionnaire for collecting ratings from job experts. An example of a task list questionnaire concerning land navigation is provided in Table 8.

The level of detail required to ensure adequate coverage depends on the nature of the application to be developed (intelligent tutors demand very detailed descriptions, development of performance tests require moderate detail, training outlines typically demand much less detail), the adequacy of existing descriptions, and the job familiarity of the job analysts/application developers. One criterion to employ is to include sufficient detail to distinguish among levels of expertise. To achieve this goal, we have found it necessary to describe each of the methods available to accomplish higher level task goals.

Table 8
A Partial List of Land Navigation Tasks

Duty *Tasks* Methods	Average Diagnosticity Ratings
Land Navigation	
Determine location	
Determine position by terrain association	3.8
Locate an unknown point by intersection	2.3
Determine position by 1 point resection	1.8
Determine position by 2 point resection	1.5
Determine distance	
Estimate ground distance visually	3.0
Determine amount of time to cover ground distance, given	2.5
Determine distance on a map	1.3
Determine number of paces to cover ground distance	1.3
Determine direction	
Preset compass under dark conditions	2.5
Determine magnetic azimuth using centerhold technique	2.3
Convert magnetic azimuths to grid	1.8
Determine magnetic azimuth using compass to check method	1.5
Determine grid azimuth	1.3
Plot grid azimuth using protractor	1.0

For example, note the three levels of detail displayed in Table 8, indicated by the three font styles (bold, italic, and plain). The most abstract level, job duties, describes similar groupings of tasks. Typically, duties are based on sharing the same overall purpose--in this case, the purpose is navigating to a point on land. The task level provides a general description of an activity to accomplish a particular goal (e.g., determine location). The task statements presented in Table 8 represent the finest level of analysis found in typical job analyses in personnel psychology.

However, discriminations among levels of expertise cannot be made at the task level. Expertise can be described by how well persons perform tasks, not which tasks they perform. The description of performance levels begins with an account of which method is used to achieve the task goal. Statements describing methods used to accomplish goals represent the third level of detail shown in Table 8. Although information about task methods can sometimes be gleaned from job documents, more often it requires additional work via interviews and protocol analyses to articulate the expertise associated with actual job performance.

A limitation of the task-list format is that it does not clearly show how the various tasks are linked and related, nor does it easily accommodate recording details about successive decompositions of task components. Such information is useful for designing curriculum, intelligent tutors, technical manuals, and other applications. To better capture information of this sort, we also employ plan-goal graphs.

Plan-Goal Graphs. Plan-goal graphs are graphical representations of task structure (Rouse, Geddes, & Hammer, 1990; Sewell & Geddes, 1990). The plan-goal graph decomposes the most abstract purpose of a task (or job) into increasingly resolved descriptions of performance, until the descriptions are sufficiently detailed and complete for the purpose of your application. Goals indicate the purpose of a plan, generally in terms of desired states of the world. A goal can be satisfied by any one of its subordinate plans. Plans specify the alternative methods available for satisfying a goal.

A portion of a plan-goal graph for computer maintenance is displayed in Figure 2. The goals are represented by ovals and the plans are shown as boxes. Thus, the "gather more data plan" and the "use timing diagram plan" constitute two of the four different methods for achieving the goal of "cause identified". The different methods are potentially disjunctive; executing any one of them will satisfy the goal. On the other hand, goals are always conjunctive. For example, note in Figure 2 the goals "relevant figure in view" and "start identified". Both of these goals must be accomplished to achieve the plan "use flowcharts". Thus, goals also provide completion criteria for plans. When all of the sub-goals under the "use flowcharts" plan are satisfied, the plan is completed, and so is its parent goal, "identify cause".

The verbal labels, the particular decomposition, and the depth of the decomposition in a plan-goal graph reflect a certain amount of discretionary decision making. Any domain can be described in a variety of ways and at different levels of abstraction, none of which is objectively more correct than the other. To help tolerate this ambiguity, it might help to realize that the plan-goal graph is only a representation of domain knowledge, in the same way that a map is a representation of the world. The fidelity of a map and even the accuracy of the locations depicted depend on the purpose of the map. For example, the location of streets on a city map is sufficiently precise to support driving decisions. But the distance between stops on a subway map often departs dramatically from their depiction on a map for driving. The purpose of these deviations are to help the rider recognize stops for transfer and departure.

When developing a plan-goal graph, one issue that occurs is knowing when to distinguish two different plans for the same goal. The criterion we use is when the candidates involve qualitatively different concepts that cannot be captured by adjusting the range of a quantitative parameter (Geddes, 1989). For example, the four different plans for determining the cause of the fault involve strategies and different features of the task environment. When two plans do share knowledge, it is indicated by having them point to the same lower-level goal and plan in the decomposition.

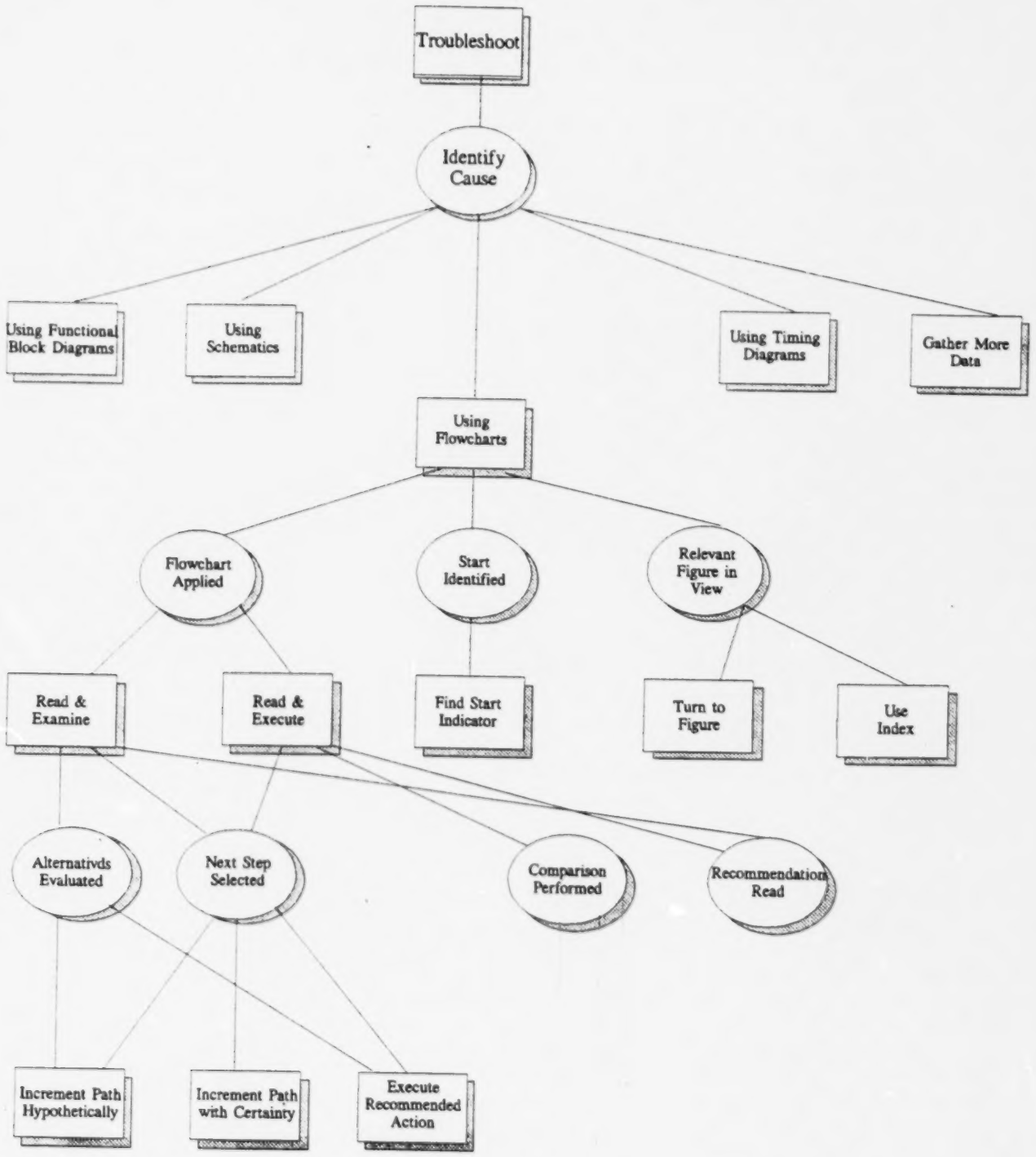

Figure 2. Portions of a plan-goal graph for the computer technician job.

BEST COPY AVAILABLE

We annotate each plan in the plan-goal graph with information required to support plan implementation, such as the declarative, procedural, generative, and self knowledge components described in our model of job expertise. We also annotate the plan-goal graph with descriptions and estimated distributions of typical mistakes.

If the purpose of your task analysis is to provide a description for developing a computational system that performs some of the same work activities, the specific domain decomposition that you generate is probably important. You'll need more guidance than we provide in this report. But if your purpose is to develop job knowledge tests or training curricula, achieving the purpose is probably fairly robust in the face of potentially many different decompositions of a domain. You should be primarily concerned with completeness and indicating task interactions.

The plan-goal graph has two advantages for applications of cognitive task analyses. First, the plan-goal graph clearly illustrates the domain-specific goal structure of performance, an important element of job expertise that is missing from task-list representations. Second, it ensures that test content is directly relevant to task performance by requiring knowledge to be explicitly linked to job goals. Further, it describes the relationships between goals and methods at several levels of detail.

Summary of Cognitively-Oriented Task Analysis

Cognitively-oriented task analysis involves a breadth then depth approach to describing job expertise. We engage job experts in interviews and questionnaires to define the tasks comprising a job, and to identify tasks that best reveal the nature of job expertise. We then employ protocol analyses of performance in context to elicit the knowledge requirements for performance.

By examining expert performance in actual work setting, the results identify knowledge that is often overlooked or ignored by conventional methods of task analyses. By systematically sampling the people, tasks, and contexts comprising job performance, the approach is comprehensive and relevant.

This task analysis approach has been successfully employed for a variety of uses in several domains. It has been applied to the domains of land navigation and computer technician performance and has been used to develop measures to predict and to diagnose performance, and to evaluate training needs. In the next section we describe how these task analysis results are used to develop written performance measures.

One limitation of this approach involves its primary reliance on protocol data. Although concurrent verbal reports reveal some contents of current awareness, many perceptual processes occur too quickly to be verbalized or are not sufficiently articulated to be spoken. While it has been successfully used for tasks requiring perceptual knowledge, its success depends on the degree to which perceptual knowledge is already articulated by job incumbents.

The selection and adaptation of task analysis methods requires attention to both organizational feasibility and scientific validity. Developing quality applications involves balancing tradeoffs. The criteria we employed for developing our cognitively-oriented approach to task analysis are shown in Table 9. These criteria provide some perspective on the choices involved in developing an appropriate task analysis strategy.

Table 9
Criteria for Selecting Task Analysis Methods

Criteria	Description
Completeness	Adequacy and appropriateness of methods for describing tasks and knowledge.
Accuracy	Fidelity to job performance. Agreement among job experts.
Cost effectiveness	Match to application requirements and organizational resources.
User acceptance	Response of users and management to task analysis processes and results.
Timeliness	Project duration and amount of personnel resources required for completion

The impact of these considerations on task analysis methods were substantial. We highlight some of the adaptations we made in Table 10 by comparing some features of cognitively-oriented task analysis to prototypical task analysis methods from personnel psychology and cognitive science. The comparisons illustrate the general features of the approach--a focus on expert performance in context; systematic sampling across people, tasks, and contexts; and the use of videotaped protocols to identify job expertise.

Table 10
A Comparison of Task Analysis Methods

Task Analysis Activity	*Cognitively-Oriented*	*Personnel Psychology*	*Cognitive Science*
Information source	Job performance	Training materials, Job description	Laboratory performance
Task description	Interviews, Task ratings	Interviews, Task ratings	Prescription by expert
Sampling method	Stratified	Random	Prototypical
Sampling basis			
People	Levels and types of expertise	Demographic variables	Levels of expertise
Tasks	Importance Performance variability	Importance Frequency	Diagnosticity for expertise
Contexts	Importance Performance variability	List all	Prototypical
Knowledge elicitation	Video protocols	Questionnaire ratings	Verbal protocols
Knowledge representation	Plan-goal graph	List of knowledge categories	Computational model

Section 3: Develop Performance Measures

The multiple choice test has been the staple of educational assessment for nearly a century. Despite this fact, it can be characterized fairly as equal parts of art and science. Although numerous statistical tools are available for identifying good questions--once they have been written and administered--scant guidance exists for writing them. Consequently, those faced with this task are required to develop a large pool of items, several times more than is actually used in the test. Where feasible, item statistics are then used to winnow the pool. Using trial and error in item writing, an effective subset of items measuring the intended content is eventually identified through item analyses.

In this section, we attempt to improve the item writing process by providing more systematic and detailed specifications to item writers. This approach is based on the theory that content is the critical key to developing effective test questions. Thus, this section will mainly emphasize what content to include in test questions and how that content should be structured.

We present a typical approach to developing tests in Table 11. It outlines the major steps of the process. In this report, we direct our discussion to steps 1 through 5, for two reasons. These steps determine the nature and usefulness of test content. They also receive considerably less attention in most books on tests and measurement.

Table 11
Test Development Process

1	Develop test plan
2	Conduct job/task analyses
3	Develop test specifications
4	Write test questions
5	Review and revise test questions
6	Conduct pilot test
7	Edit and select items for test
8	Administer test
9	Score test
10	Validate decisions using test scores

Specifying Test Content

Identifying relevant content involves two major tasks--specifying the relevant job knowledge and defining how it will be sampled for the test (tasks 2 and 3 in Table 11). To assist item construction, we organize information from the task analysis into a tabular format at three levels of analyses. These levels are categories, tasks and methods, and knowledge elements. This representation of job knowledge follows the model of job expertise presented in Tables 1and 2 of this report. At the most general level, task analysis results are organized into categories of task and knowledge requirements.

Table 12
Description of Land Navigation Expertise

Task Categories	*Knowledge Components* A Principles/ Concepts	B Procedure Selection	C Procedure Execution	D Goal Knowledge	E Pattern Recognition	Row Totals
1 Planning	4	3	6	3	4	20
2 Location	6	4	8	4	6	28
3 Distance	4	2	4	2	4	16
4 Direction	4	2	4	2	4	16
5 Moving	4	4	6	2	4	20
Column Totals	22	15	28	13	22	100

Note: Numbers represent percentage of the total number of test questions.

To illustrate the application of task analysis results to performance measurement, we use data from our land navigation study. An example of the description of expertise for land navigation at the category level is presented in Table 12 (this category level of analysis was displayed previously in Table 1 for a general model of expertise and Table 7 for computer technicians). The numbers in Table 12 reflect experts' judgments about the relative contribution of each category to a description of the nature of land navigation expertise. For the task analysis phase, this information provided the basis for a sampling plan to target knowledge elicitation efforts. For the development of performance measures, we use this same information to to representatively sample job knowledge for a written test. From this standpoint, the numbers in Table 12 can be interpreted in terms of percentage of test content. For example, Table 12 specifies that 28% of test content should address the task of determining your location.

To be more useful to test designers, we need to provide test specifications that are more detailed than those provided by the categories of Table 12. At the next level of detail, we list the tasks and methods employed for accomplishing each category of tasks (displayed previously in Table 8). Recall from the task analysis section that these tasks and methods were also rated for their contribution to describing job expertise.

At the most detailed level, we describe the elements of knowledge required to support performance of each method. The example provided in Table 13 displays the steps for executing three methods of determining location, with their associated knowledge requirements of concepts, procedure selection, goal knowledge, and pattern recognition. Additionally, at this level of analysis, we also present information about the types and frequencies of errors that typically are made when performing this method. Information at this level most directly supports the writing of test questions. Using the information from the category and task/method levels, we can now develop a more detailed set of test specifications to guide the selection of questions for writing.

Selecting Questions to Write

Once job expertise has been clearly defined, the next task is to specify a plan for sampling this content in your test. There are few times when a subject, task, or job can be assessed exhaustively. Even rather simple workplace tasks require a surprising amount of information to support competent performance. Thus, selecting which questions to write is a critical element of effective test development and test use. Your test plan should specify a goal for the total number of test questions to write, and provide a breakdown of this total into goals for tasks and knowledge requirements.

At a general level, the model of job specific expertise accomplishes this objective (i.e., Table 12). In this table, the sampling plan is specified as a percentage of total test questions for each cell in a matrix of tasks and knowledge. For example, this model of land navigation knowledge indicates that twenty-eight percent of test content should focus on the task of determining location, with six percent of test questions addressing the pattern recognition aspects of this task.

By following this plan, test content should proportionately reflect the expertise required for effective performance of the job. This model of job expertise provides a plan for systematically sampling all areas of job knowledge according to their importance to the job and their usefulness in distinguishing levels of expertise. While it provides specific goals for each category of content, more detailed information is needed to assist writing of test questions.

Table 13
Portion of a Method by Knowledge Element Matrix

Job: Marine Corps Infantryman Duty: Land Navigation Task: Determine location Model Level: Knowledge Element	Method		
Category Element	Determine Position By Terrain Association	Determine Position By One Point Resection	Determine Position By Intersection
Procedure Execution			
Orient the Map	X	X	X
Scan the ground	X	X	X
Identify the major & unique features	X	X	X
Compare shape, size, orientation, slope	X		
Determine magnetic azimuth		X	X
Convert to back azimuth		X	
Convert to grid azimuth		X	X
Plot azimuth		X	X
Move to identifiable location			X
Determine grid coordinates	X	X	X
Goal Knowledge			
Read coordinates at center point		X	X
Confirm location using 3+ features	X		
Pattern Recognition			
Must identify recognizable features	X	X	X
Map symbols, legend info		X	X
Terrain features on ground	X	X	X
Terrain features on map	X	X	X
Procedure Selection			
Select location finding method	X	X	X
Select major, unique features	X		
Concepts & Principles			
Properties of identifiable location	X	X	X
Grid representation of geography		X	X
Grid & Magnetic azimuths		X	X
Errors			
Procedural			
Missing step		10%	20%
Insufficient precision	30%	10%	10%
Feature misidentified	45%	5%	5%
Incorrect azimuths		10%	20%
Grid coordinates misread	10%	10%	10%
Computational (math errors)		15%	20%
Strategic			
Ineffective plan	5%	5%	5%
Tactical			
Inefficient method	5%	5%	
Poor steering mark, feature		20%	
Conceptual			
Magnetic, grid & true north	5%	10%	10%

We utilize both the category (Table 12) and task/method (Table 8) levels of our model of expertise to develop a more detailed set of test specifications. We use the ratings from the task list to allocate the percentage of test questions across the set of methods for each task. We employ a top to bottom sampling strategy to meet the goal specified by the model in Table 14. For example, Table 14 shows that we need to write 28 items for the task of determining location. Using the ratings of task diagnosticity obtained from job experts (Table 8), we distribute these 28 questions across the four different methods for determining location.

Table 14
Detailed Test Specifications

		Sampling Over Knowledge Requirements				
Duty *Tasks* Methods	*Sampling Over Methods*	Principles Concepts	Procedure Selection	Procedure Execution	Goal Knowledge	Pattern Recognition
Land Navigation						
Determine location	*28*	*6*	*4*	*8*	*4*	*6*
1 Terrain association	*12*	3	1	2	2	4
2 Intersection	*7*	2	1	2	1	1
3 One point resection	*5*	1	1	2	1	0
4 Two point resection	*4*	0	1	2	0	1

Next, the test questions are distributed across each knowledge requirement so that the marginal values are maintained for each method (see Table 14). Ideally, this is done by a set of job experts, to enhance judgment reliability and accuracy. However, this task was done by a single job expert in our land navigation example owing to a limited pool of job experts.

While this task can be computed mechanically using the values in the row and column margins (in italics), more useful values can be obtained by utilizing a job expert's judgment. For example, examine the values assigned to the four methods of determining location. Using only the marginal values, more questions should be assigned to the procedure execution cell for terrain association and fewer to pattern recognition. However, job experts know that competent performance of this method requires a substantial amount of pattern recognition. Additionally, the procedural elements of this method overlap with other methods, so those portions can be assessed with questions assigned to other methods.

The marginal values represent judgments averaged over the entire domain. Hence, adjusting the values to each method should improve the fidelity and job relatedness of the test to job performance. Hence, the final distribution of test questions in this detailed test specifications reflect both the detailed knowledge requirements for each method, and the overlap in knowledge requirements between methods (e.g., methods sharing some of the same procedural steps or concepts).

In sum, we employed task analyses results to provide detailed test specifications in a task by knowledge format. As a result the performance measure should proportionately reflect job expertise. This approach accomplishes two objectives. First, it provides clear guidance to the test developer by specifying which methods and knowledge requirements should be assessed. Second, it ensures that test content reflects job content by representatively sampling both tasks and knowledge.

Writing Items

The major point of this section is to improve test development and test quality by more clearly specifying what test content should be. By using a cognitively-oriented approach to task analysis, these objectives have been accomplished by identifying the tasks, methods, and knowledge requirements that experts employ when performing the job. In particular, considerable attention has been paid to specifying the knowledge requirements in some detail. Consider the following question from an existing test of land navigation. It assesses knowledge related to the task of determining direction.

> 1. To measure an azimuth, you look through a rear sight notch and align the sights by centering the front sight hairline in the rear sight notch. What technique are you using to determine this magnetic azimuth?
>
> a. Compass-to-cheek technique
> b. Recon technique
> c. Compass-point technique
> d. Centerhold technique

This question assesses an examinee's knowledge of a fact, the name of a direction-finding procedure. Although it is probably true that most good navigators know the correct answer is "a", this fact is incidental to effective performance of land navigation. To determine your direction using this procedure, you ordinarily will not need to use its proper name.

An important rule for writing effective test questions is to frame the question so that the examinee will process information in the same way as is done on the job. By using the task analysis results (Table 13, column 3, 'Determine Position By Intersection'), we constructed a question which assesses the same land navigation task, but requires the examinee to employ his knowledge in the same way as would be done on the job. Additionally, we framed the question in a realistic scenario drawn from performance examples gathered during the task analysis.

> 2. You are a security outpost for your patrol in a hostile country. Your patrol is located on the hilltop at grid coordinate 016726. Looking to the southwest, you see an enemy patrol stopped along a secondary hard road. Using your compass, you determine that the magnetic azimuth to their location is 237.5 degrees. To identify the enemy location to your command, what 6 digit grid coordinates will you report?
>
> a. 738983 *procedural error*
> b. 981736 ***correct response***
> c. 983738 *procedural error*
> d. 736981 *procedural error*

This question requires the examinee to perform the task of using direction information to determine the position of a distant location. To answer this question correctly, the examinee must perform the same

operations, and in the same manner, as he would for his job of Marine infantryman. That is, he must first correctly locate his own position on the map, given the grid coordinates stated in the question. Then he must convert the magnetic azimuth to a grid azimuth and precisely plot it on the map. Finally, he must correctly read the 6 digit grid coordinates of the intersection of the plotted azimuth and the road.

This question assesses an examinee's understanding of procedural knowledge that is required for competent performance. The previous question assesses declarative knowledge that is related to, but is not required for job performance. In a following section, we will describe how using questions that are directly relevant and essential to performance improves the validity of measurement.

In comparing the questions, two additional features should be noted. The response alternatives to the second question represent answers that would be given if common errors are made in performing the procedures. For example, the first and last answers result from reading the map coordinates in the wrong order--a mistake often made by novices. Response alternative 'c' results when examinees fail to convert the azimuth from magnetic to grid. Thus, even wrong responses provide useful information for diagnosing and predicting examinee performance.

By comparison, two of the responses for the first question were entirely made up. The other incorrect response can be ruled out by savvy test takers from information in the question stem, without any knowledge of land navigation. Consequently, both correct and incorrect answers to this question have multiple, and ambiguous, interpretations. This ambiguity reduces the validity and the interpretability of test scores. In contrast, information about even incorrect responses potentially can contribute to both diagnostic efficiency and predictive validity. We also suspect that it may contribute to examinee perceptions of test fairness and validity.

A second useful feature of the second question is that it is framed in a realistic scenario. This may help maintain examinees' interest and acceptance of the exam. Importantly, it may also help to motivate the examinee to learn and remember the information presented, by demonstrating how it will be used and suggesting some of the consequences of not knowing it.

Question Stems

The example questions underscore our theme that content is a primary contributing factor to test quality. In recent years, the trend has been for tests to include more questions assessing procedural, rather than declarative knowledge. For tests with goals of assessing job performance, this shift will result in improved validity of assessment, diagnosis, and prediction.

However, the assessment of procedural knowledge tends to be limited to testing <u>how</u> procedures are performed. Other aspects of procedural knowledge are also essential to support competent performance on the job. As presented in our previous discussion of a model of job expertise, these include knowing when to use a procedure (procedure selection), knowing what standard of precision is required, and recognizing perceptual patterns that guide task performance.

By more precisely specifying the knowledge requirements of performance, clear guidance is provided to item writers for designing the body, or stem, of test questions. In this way, the question stem is constrained by the cognitively-oriented task analysis to specific methods and knowledge requirements, relevant to competent job performance. Thus, we can use the framework of knowledge requirements (see Table 2) as a taxonomy of question types. We illustrate this point with some examples from our applications in land navigation and computer maintenance.

Procedural Knowledge/Procedure Selection. Questions addressing this knowledge requirement assess examinees' skill in deciding which of several available methods should be employed in a given situation. The key to writing good questions of this type is to adequately capture some of the complexity and ambiguity in situations which realistically occur on the job.

Typical of procedure selection questions in many domains, none of the answers in the following example are actually wrong. However, one response provides a substantially better result in terms of both speed and accuracy. Selecting the optimal response requires matching characteristics of the situation with the conditions and constraints for implementing each method.

> 3. PVT Rojas is following an azimuth of 166° to a checkpoint 1200 meters from his start point. He has moved 600 meters through a forest, and believes he may have drifted off course while weaving through the trees. From his map, he sees that the last 400 meters of the route goes through a clearing with road across his path at 1000 meters. He scans the immediate area but can't see far because of the trees. What should he do to get back on course?
>
> a. Return to the start point and begin again
> b. Recon the area and plan a new route
> c. Continue on his azimuth until the road, then adjust
> d. Perform resection to determine his current position

In this example, it requires knowing what the resource and time requirements of each method are. Response 'c' produces a much more efficient result. Responses 'a' and 'b' require too much time. Performing response 'd' requires visually locating major and unique features which are identifiable on the map. Their current location in a forest makes this method difficult to implement.

Procedural Knowledge/Goal Understanding. Questions that assess goal knowledge address whether examinees know when a procedure is complete, what standard of precision is required, and what are the relative priority of competing goals.

> 4. Standing on Smith Road, you determine that the grid azimuth to Crowder hill is 335°. From your map, what is the 6 digit grid coordinates of your current location?
>
> a. 506917
> b. 507919
> c. 506918
> d. 507917

This example assesses precision. In order to select the correct response, the examinee must both plot an azimuth and read map coordinates with adequate precision. The use of an unsharpened pencil or careless placement of the protractor could result in errors of 200 meters or more. For Marine infantry, errors of this magnitude could lead to potentially serious consequences, such as running into a minefield.

Procedural Knowledge/Perceptual Information. Competent task performance often requires perceiving and interpreting visual cues correctly. This may be required to support the choice of a method, performance of procedural steps, or recognition of a problem or change in status. Sometimes this perceptual knowledge involves identifying relevant cues out of complex stimuli, while for other tasks it involves recognizing

patterns of cues. In the following example, it involves interpreting contour lines on a map which would be provided to examinees for the test.

> 5. While planning a route to your next checkpoint, you need to evaluate which hills can be more easily traversed by foot. Which of the following best describes the slope on Map A from grid coordinates 938745 to 942745?
>
> a. Steep downward slope
> b. Gentle downward slope
> c. Steep upward slope
> d. Gentle upward slope

Declarative Knowledge/Concepts. Typical of many written multiple choice exams, the first example in this section (question 1) assessed declarative knowledge. We criticized this question because it assessed information that was not essential to task performance. This criticism is directed at the relevance of the content rather than its nature (i.e., declarative knowledge). The next example assesses declarative knowledge that is important to land navigation--the properties of steering marks used to keep navigation on course.

> 6. Corporal Johnson is navigator for a team moving through unfamiliar territory. There are several easily distinguished objects along their line of march that he could use for steering marks. Which quality should most affect his decision?
>
> a. Brightness
> b. Height
> c. Nearness
> d. Distance

Response Alternatives

The knowledge element table (Table 13) is also used to generate response alternatives. The bottom portion of the table displays the type and distribution of errors that are typically made when performing each of the methods listed. The errors are classified into one of several types, based on the content of the mistake. These errors directly correspond to the knowledge requirements displayed in the upper portion of the table. That is, procedural errors correspond to mistakes in procedure execution and so forth. Computational errors are one type of procedural error that was identified to increase diagnostic efficiency. Similarly, strategic and tactical decision errors correspond to procedure selection. These differ with respect to whether the decision difficulty involves errors in planning or errors in adjusting plans during implementation to specific situations.

Using this task analysis information provides several advantages to item writers. First, it provides a variety of choices for creating a set of response options. Because each error actually occurred on the job, it also ensures that the response options are plausible. Further, several useful rationales for selecting among the choices can be devised using the task analysis information. For example, when the purpose of the test is diagnostic, each question's response options can be constrained to one type of error to increase diagnostic efficiency. We employed this strategy for the example questions previously presented. However, if the purpose is to predict performance, then response options can be chosen across all classes of errors using the error distribution information in the table to select the most frequently occurring errors.

When response options are structured in this way, then scales can be constructed using information from incorrect as well as correct responses. For example, we constructed a scale for the land navigation test that consisted of incorrect responses based on computational errors. This scale could then be used to identify individuals who specifically needed tutoring in math to improve their land navigation skills. It is also possible that information from incorrect responses can improve the predictive efficiency of test scores. Thus, performance predictions may vary for individuals with the same test score, based on the nature of their respective errors. It is possible to easily recover from some procedural errors, while strategic and tactical decision errors are usually more costly.

Reliability and Validity

One of the key strategies of the cognitively-oriented approach to task analyses has been to utilize the judgments of job experts. The primary advantage for doing so is the efficiency gained by targeting the use of task analysis resources. The reliability and accuracy of their judgments has been an area of some concern for us in assessing the tradeoffs between gains in efficiency and losses in fidelity. The issue is that even experts are frequently unaware of their own, much less others', knowledge and cognitive processes.

To address this issue, we gathered data on several of the judgment tasks where we involved job experts. The results for three judgment tasks are presented in Table 15. The first task addressed the relative contribution of each of the categories of tasks and knowledge to job expertise (i.e., Table 12). The second task involved rating the diagnosticity of methods within each task (see Table 8). The third task consisted of estimating the relevance, proportion correct, and item-test correlation for land navigation test questions. The judgments were made independently and each task involved a different set of job experts. The inter-rater reliability among judges ranged from .71 to .86 for these judgment tasks, indicating an acceptable level of agreement.

Table 15
Reliability and Validity of Expert Judgments

Judgment Task	Dimension Rated	N of Stimuli	N of Raters	Inter-rater Reliability	Validity
Job Duties	Diagnosticity	10	5	.86	.65**
Tasks	Diagnosticity	63	4	.78	-.24*
Test Questions	Proportion Correct	65	3	.78	.56**
	Diagnosticity	65	3	.73	-.18
	Relevance	65	3	.71	.33**

Notes:
1. *p<.05, **p<.01
2. 'Validity' is the correlation between mean ratings of judges and a relevant empirical index. For each task, the empirical index was the average of item-criterion correlations.

To estimate the accuracy of experts' judgments, we correlated the mean judgments for each stimulus in each task with a corresponding index, estimated empirically from test data. The index used for each judgment task consisted of mean item-criterion correlations computed for each test question. Each test question had previously been classified according to which task and task category it addressed. For the first task, the mean item-criterion correlation was computed for each of the 10 categories of tasks and knowledge. These indices were then correlated with the judgment means from the job experts. Similarly, mean item-criterion correlations were computed for each task and correlated with the corresponding mean from job experts' judgments. For the third task, the rationally estimated item indices were correlated with corresponding empirically estimated indices.

The validity results are mixed. At the category level, experts' judgments correlated well with the mean item-criterion correlations, suggesting that job experts can make meaningful judgments at this general level of analysis. However, at the task level, the correlation is actually negative. Similarly, diagnosticity ratings made at the item level were also negatively correlated with their empirical counterpart--item-test correlations. After carefully reviewing the judges ratings, one possible interpretation is that judges tended to confuse diagnosticity with difficulty. If true, this suggests that rating instructions need to be improved, with an additional study to confirm this interpretation. Finally, results for judging the content relevance and proportion correct were significantly correlated with empirical item indices. Overall, the results indicate that job experts can make meaningful judgments, but that their understanding of rating 'diagnosticity' is suspect.

Table 16
Content Analysis of Land Navigation Tests
(In percentages)

Content Categories	Existing Land Navigation Tests							Cognitively-Oriented Test
	1	2	3	4	5	6	Avg	
Tasks								
Planning	0	0	10	0	0	0	2	17
Location	34	56	42	66	62	72	55	38
Distance	6	18	15	22	21	28	18	16
Direction	54	26	31	12	17	0	23	16
Movement	6	0	2	0	0	0	1	13
Knowledge								
Principles/Concepts	7	0	6	0	0	0	2	8
Procedure Selection	0	0	3	0	0	0	1	12
Procedure Execution	33	55	25	67	67	46	49	32
Goal Knowledge	0	0	0	0	0	0	0	4
Pattern Recognition	0	42	35	33	29	18	26	38
Declarative Knowledge	60	3	31	0	4	36	22	6
Test Length	15	30	128	9	24	11	36	100

Note: Numbers represent percentage of test content

Next, we examined the content of all existing land navigation tests that we could locate. Using the categories we developed in our task analysis, two members of our research team independently rated the content of each item of each test. Each item was classified into one of the five task categories, then one of the five knowledge categories. Comparisons of content between six existing tests and the cognitively-oriented one we developed are exhibited in Table 16. Differences in content are clear. Existing tests give little attention to the two task categories that emphasize decision-making--planning and movement. These results are consistent with what would be expected of task analyses that fail to adequately capture the mental aspects of performance. Similarly, existing tests substantially under-represent knowledge content related to principles, procedure selection, and goal knowledge.

The next question is to determine whether these differences in test content, presumably due to their respective task analyses, are related to differences in validity of measurement. To address this question, we compare the correlations of the knowledge test to two measures of performance, hands-on measures of proficiency and integrated performance tests assessing navigation to four checkpoints in a wilderness setting.

The results of these comparisons are shown in Table 17. The first two rows display a direct comparison between an existing land navigation and the cognitively-oriented one. One group of subjects from our study had recently been assessed using existing measures of both written and performance tests, then were given the experimental written and performance tests one month later. The cognitively-oriented test significantly outperformed the existing measure for both performance measures.

Next, we compared the correlation of the cognitively-oriented test with hands-on measures of skill to all other job knowledge--hands-on test correlations we could locate in the scientific and technical literature. Again, the results indicate that the cognitively-oriented measure better corresponds to hands-on measures of performance. These results suggest that the additional categories of content included in the cognitively-oriented test are important to competent land navigation performance. By extension, these results also imply that the cognitively-oriented task analyses identify knowledge essential to performance which are missed by existing procedures.

Table 17
Correlations of Job Knowledge and Performance Measures

		Performance Tests		Hands-on Skill Tests	
Test	N	1	2	Observed	Corrected
Cognitively-oriented Landnav	31	.51	.48		
Existing Marine Landnav	31	.51	.08		
Cognitively-oriented Landnav	358			.58	.72
Summary from scientific literature	11,949			.41	.59

Conclusions

There is clearly a practical need for applied cognitive task analyses to support the development of applications for training, measuring, and improving performance. Recent improvements in task analysis focus on the capability of identifying what individuals learn from job experience. The challenge in this task is the complexity of workplace performance. Job expertise is simply not unidimensional. It encompasses competence in technical, interpersonal, perceptual, and motor dimensions of performance, across a wide variety of tasks and contexts. Further, there often are several ways to perform competently.

To meet these multiple challenges, we integrated the concerns, content, and methods of personnel psychology and cognitive science. Personnel psychology has long been concerned about issues of the dimensional structure of job performance, sampling and generalizability across persons, tasks, and contexts. Cognitive science has focused on specifying in detail the nature and content of task expertise. Capturing the essential content of *job* expertise requires the contributions of both. We utilized methods from personnel psychology to describe the breadth of job tasks and methods from cognitive science to identify the depth of job knowledge. From our work, it also appears that the whole of job expertise is greater than the sum of its parts. Our task analysis work reveals that much of what has been missing using existing task analysis methods is the mental aspects of performance related to interactions among task dimensions, task characteristics, and contexts.

Acknowledgements

This work was supported by the Office of Naval Research, Cognitive Sciences Program and the Manpower, Personnel, and Training Program under contract N00014-93-C-0199. We appreciate the encouragement and support of Dr. Susan Chipman. Thanks are also due to LCDR Don Sine (now retired), Gary Mahler, and FCC Chris Haden of the AEGIS Training Center for their support and assistance in studying expertise associated with computer technicians.

References

Campbell, J. P. (1990). Modeling the performance prediction problem. In M. D. Dunnette and L. Hough, (Eds.), *Handbook of Industrial and Organizational Psychology* (pp. 687-732, 2nd ed., vol. 1). Palo Alto, CA: Consulting Psychologists Press.

Campbell, J. P., McCloy, R. M., Oppler, S. H. & Sager, C. E. (1993). A theory of performance. In *Personnel Selection in Organizations*, N. Schmitt and W. C. Borman (Eds.). San Francisco, CA: Jossey-Bass, Inc.

DuBois, D., & Shalin, V. (1995). Adapting cognitive methods to real world objectives: An application to job knowledge testing. In P. Nichols, S. Chipman, and R. Brennan (Eds.), *Cognitively Diagnostic Assessment*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Dunnette, M. D. (1963). A note on the criterion. *Journal of Applied Psychology*, 47, 251-254.

Ericsson, K. A. & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin, 51*, 327-358.

Gagne, R. M., Briggs, L. J., & Wager, W. W. (1988). *Principles of instructional design* (3rd ed.). New York: Holt, Rinehart, & Winston.

Gelman, R. & Gallistel, R. (1978). *The child's understanding of number*. Cambridge, MA: Harvard University Press.

Geddes, N. D. (1989). Understanding human operator's intentions in complex systems. Unpublished doctoral dissertation. Atlanta, GA: Georgia Institute of Technology.

Klein, G., Calderwood, R., & MacGregor, D. (1989). Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man, and Cybernetics, 19*, 3, 462-472.

Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology, 78*, 2, 311-328.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231-259.

Orasanu & Fischer, (1992). Team cognition in the cockpit: Linguistic control of shared problem solving. In Proceedings of the 14th Annual Meeting of the Cognitive Science Society. Hillsdale, NJ: Lawrence Erlbaum.

Rouse, W. B., Geddes, N. & Hammer, J. M. (1990). *Computer aided fighter pilots*. IEEE Spectrum, March, 38-41.

Sewell, D. R. & Geddes, N. (1990). A plan and goal based method for computer-human system design. Human Computer Interaction, INTERACT 90, New York, NY: North-Holland. (pp.283-288).

Smith, P. C. & Kendall, L. M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology, 47*, 149-155.

Appendix A:
An Example of Knowledge Elicitation and Representation

Table 1 provides an excerpt from our knowledge elicitation activities in the domain of electronics repair. We use this to illustrate how we apply our task analysis suggestions. The table indicates the speakers in the leftmost column, which includes numbers for later reference. The transcription of their verbalizations is provided in the center column. The rightmost column contains our interpretations of the speaker's verbalizations.

The session from which this illustration was drawn included two observers and two instructors, one is a navy chief and the other is a civilian instructor. The session occurred in a land-based laboratory used for instructional purposes. The equipment in this laboratory closely resembled the computer room of ships, but also included capabilities for inserting simulated faults. The civilian's role in the session was to select and program faults into the equipment, and to discuss alternative faults with the observers.

As a subject matter expert (SME), the chief's role was to conduct his ordinary diagnostic activities while thinking aloud. We expected him to be very good. However, his current job assignment involves teaching and administrative work. The recent absence of frequent and challenging hands-on work creates the possibility that the chief is a "decayed expert". This category of expert retains all of the conceptual aspects of domain knowledge but loses some ability to apply this knowledge in specific situations. The excerpt begins in the middle of the chief's attempt to localize a fault. He is just completing a test with the voltmeter, with some assistance from the civilian instructor.

We chose this excerpt for several reasons. First, the excerpt illustrates the challenges of knowledge elicitation: a) the SME was not comfortable thinking-aloud and required numerous prompts from the observer and b) we had to manage the civilian to prevent him from embarrassing the SME. Second, the protocol provides interesting content for constructing a plan-goal graph: a) the SME worked on a problem that was not immediately obvious to him, and required substantial reasoning , b) the SME illustrates several different methods for troubleshooting and c) in several cases the SME criticizes and overrides the documentation. Finally, the protocol provides a suitable foundation for the development of questions for a job knowledge test. Following the protocol excerpt, we discuss our interpretation of this data and then present how we represented it in a plan-goal graph.

Table 1
A Protocol Excerpt From Electronic Diagnosis

(1)	Civilian: Reading out less than 1 volt. Now it reads about 4.3 volts.	The civilian reads off the value of a meter. This description will appear as part of the plan for applying the volt meter method.

(2) Chief:	Ok, so that's good.	The informant provides an interpretation of the value. This also will appear as part of the plan for using a volt meter.
(3) Civilian:	See the voltage change. Right now its uncovered. Low should be low, under one volt. As soon as I put my finger over the light, it goes up. That means the sensor is working.	The civilian is tutoring the observer. This is potentially interesting, but within earshot of the informant. Also, if the informant should feel that the observer is otherwise preoccupied, the informant may stop thinking-aloud, reducing the record we obtain for his reasoning.
(4) Obsvr:	So that's not our problem.	The observer acknowledges the civilians comment, but doesn't ask a question or encourage further comments.
(5) Civilian:	Nope.	
(6) Obsvr:	Ok.	The observer is still not encouraging further comment from the civilian. The interest here lies in the informants' verbalizations.
(7) Civilian:	Chief Smith isn't giving very positive answers today.	The civilian criticizes the informants' performance.
(8) Obsvr:	He's doing great actually.	The observer tells the civilian and the chief that she approves of the chief's performance despite the fact that he does not identify faults immediately.
(9) Chief:	I'm gonna say it stops soon after being picked up *(part of D on 5-17)*. Replace auto-thread module A-9.	The SME is in the process of using a flowchart as a method for identifying the cause of a fault. The chief assigns an interpretation to his observations. This is one indication of the challenge of interpreting observations according to domain terminology.
(10) Civilian:	Do you wanna replace it?	

(11) Chief:	Hold on, I don't want to replace anything yet. Ok. Problem still exists. There's something you can check! Let's go in and look at that. It comes down here and tells you to replace the A-9 module. And then it comes down here and tells you more places to go. To me, it would make more sense to go down here. It's silly.	This comment reveals a preference for gathering more information by conducting more observations before swapping faulty parts, despite the instructions in the flow chart. The chief refers to the documentation as "silly", perhaps suggesting a concern for efficiency.
(12) Obsvr:	Ok, so you are just gonna make a little change there on this flow chart.	The observer acknowledges the SME's departure from recommended procedures, because this might not be apparent in the video.
(13) Civilian:	Making a technician change. That's good.	The civilian changes his assessment of the chief.
(14) Chief:	Looking for the THS light. The red one right here. Now me, I don't think it is gonna light.	The informant states the purpose of his action, and points out the object of interest and states his expectations.
(15) Obsvr:	You don't think this is the problem?	The observer isn't quite sure what the chief means, but echoes a response to indicate her attention.
(16) Chief:	It's not gonna light.	
(17) Obsvr:	Oh, you don't think its gonna light. What happened?	The observer echoes the chief's comment again. She prompts him, probably because he seemed to pause too long before speaking.
(18) Chief:	It didn't light. No. Replace the tape threaded sensor. Now that would be. That doesn't make sense though. Tape threaded sensor. Why would that cause that problem. Why would that cause that problem. I've got to think about this.	The chief provides a response to the observer's prompt. The informant reveals a preference for reasoning before doing. He also warns the experimenter that he doesn't have a ready explanation and that this will take some time.
(19) Obsvr:	You don't see how it could cause that problem?	The observer acknowledges his difficulty but won't let the SME remain silent while he solves the problem.
(20) Chief:	No.	The chief treats the prompt as a question that could be satisfied with short, non-substantive answer.

(21) Obsvr: What does the threaded tape sensor do?	The observer tries a more substantive prompt that cannot be satisfied with a one word reply.
(22) Chief: It's saying that the tape is. I'm trying to think of when that tape threaded sensor light comes on. I don't know when it comes on. Where's our little time chart? I'm trying to think when it comes on.	The SME complies with the obligation to reply, and fortunately begins to verbalize on his own again. The absence of verbalization in the past few turns leaves us only with the idea that the chief is pondering a difficult problem. We have no idea of his reasoning during the silence.
(23) Obsvr: Ok, this is page 3-71.	The observer records the page number that the SME has accessed so that it may be consulted later for interpreting his following comments.
(24) Chief: Turn on, vacuumed sensed.	The SME begins to read the timing chart and then stops verbalizing.
(25) Obsvr: What are you looking at there?	The observer prompts the SME again.
(26) Chief: I'm just looking at where that sensor comes into play *(points to bottom of page, also may be looking at 3-70 or 3-69).*	
(27) Obsvr: Uh huh.	The observer provides a benign prompt to indicate her continued attention and expectations for continued verbalization.
(28) Chief: *(pause)* Set thread failure.	
(29) Obsvr: 3-69.	The observer records the page change.
(30) Chief: Counterclockwise. Clockwise. Clockwise. That's gonna send that tape across the blower sensor. So, that's occurring. Tape cross lower sensor, that makes the machine reel turn clockwise. Let's see if they turn two different circuits on it. Tape cross lower sensor	The SME finally provides an interpretation of the text.
(31) Obsvr: 5-163.	The observer records a page change.

(32) Chief:	I'm not there yet. This is the auto thread board. I'm looking where that threaded sensor comes in. It's in right here *(pause)*. It's gotta be. Here comes	The SME indicates his awareness of the observers' task by correcting her.
(33) Obsvr:	5-162.	The observer records a page change.
(34) Chief:	Its gonna come in. That doesn't make sense. Lower sensor comes in boom boom boom boom boom. *(pointing to bottom right corner)* 3 Bravo *(looking now at middle left)*.	
(35) Obsvr:	So what are you thinking of there?	The observer has remained silent as long as the SME was thinking-aloud. But she prompts the SME after a certain period of silence has elapsed.
(36) Chief:	I'm trying to figure out how that threading sensor comes into play in all this. I can see that it's probably a problem. Threaded sensor comes in and makes that turn clockwise. And then the threaded sensor is not there within 5 seconds it's gonna shut down. Which it does. My concern is what makes that threaded sensor turn off. I know what makes the threaded sensor turn on! This right here is where the threaded sensor turns on.	
(37) Obsvr:	Why are you concerned about that at all. It doesn't turn on.	The SME's pause invites a comment. The observer says something to indicate her attention and prompt more thinking aloud.
(38) Chief:	It's supposed to give you that the tape is threaded. That sensor works because things are turning clockwise. Something is goofed up, lets say that this is broken. It has to have some way to remember that positive control of the tape. If it doesn't pull tight over this hole, there's something wrong right here, let's shut it down before we get tape all over. That's why this threaded sensor is not working. That threaded sensor is your *(pause)*. Yeah, I'm looking for all the sensor. It's back here somewhere.	The SME offers a substantive reply to the query. The reply is interesting and communicates the SMEs understanding of the mechanism in question. However, because the reply is a response to an observer's query we cannot assume that this reasoning would have been part of his diagnostic processes without the prompt.

As long as the SME is fairly verbal, the observer indicates her engagement with "uh-huh" (27). The observer's comments largely indicate her continued attention, generally by paraphrasing or responding to the SMEs most recent comment (12, 15, 17). The observer becomes intrusive when the periods of silence increase (19, 21, 25). The silence is associated with complex reasoning, which is exactly the place we'd like to get the most verbalization. In all of these cases the questions and comments are not intended to elicit specific information, but rather indicate sufficient engagement to require continued verbalization from the SME. In this excerpt, the most intrusive intervention from the observer is a substantive question (37). The SME offers a substantive reply. However, the connection of the verbal response to his diagnostic reasoning is uncertain. For this reason, we minimize the use of this kind of intervention.

The civilian instructor also had the potential to influence the SME. If the civilian was actually participating as part of a diagnostic team, his influence would be part of a typical task setting and we would treat him as another SME. But in this case, the civilian is really just a third observer, one who was far better informed than the other observers, and in a good position to embarrass the SME. First the civilian attempted to engage the observer (3), who responds in a manner that closes down further conversation. Then the civilian offers a critical commentary on the SME's performance (7). Since the SME was experiencing some difficulty associated with the task, and he tended to avoid verbalization anyway, the observer wanted to support the SME and discourage the civilian from such comments (8).

In several places the SME uses documentation. The observer notes the page numbers for later reference (23, 29, 31, 33). The SME shows his awareness of the observer's goals by correcting a faulty page reference (32).

We infer that the goal of the SME's activity is to find the cause of an observed failure. The support for this inference comes from statements like (4), in which the SME suggests that he has not yet found the problem. Note that the goal here is not simply a state of the world, but a state of the SME's mind. If he believed he had found the failure, we would not expect any further diagnostic activities. States of mind are not necessarily goals. For example, item (9) is not something that the SME is trying to achieve.

The protocol indicates several methods for finding this failure. One method is to follow the instructions in a fault-isolation flow chart (9) (see Figure 1). A second method is to examine a timing diagram to determine the sequence and duration of events that should occur (22) (see Figure 2). A third method involves a functional block diagram (24) (see Figure 3). A fourth method involves a schematic diagram (34) (see Figure 4).

We organize the present problem solving in terms of four different methods, defined by the four different representations used (see four methods under goal "a" in plan-goal graph). We could have grouped the four methods as one method, perhaps called "trace diagram". The trace diagram method would have slight variations that depended on the particular diagram in use.

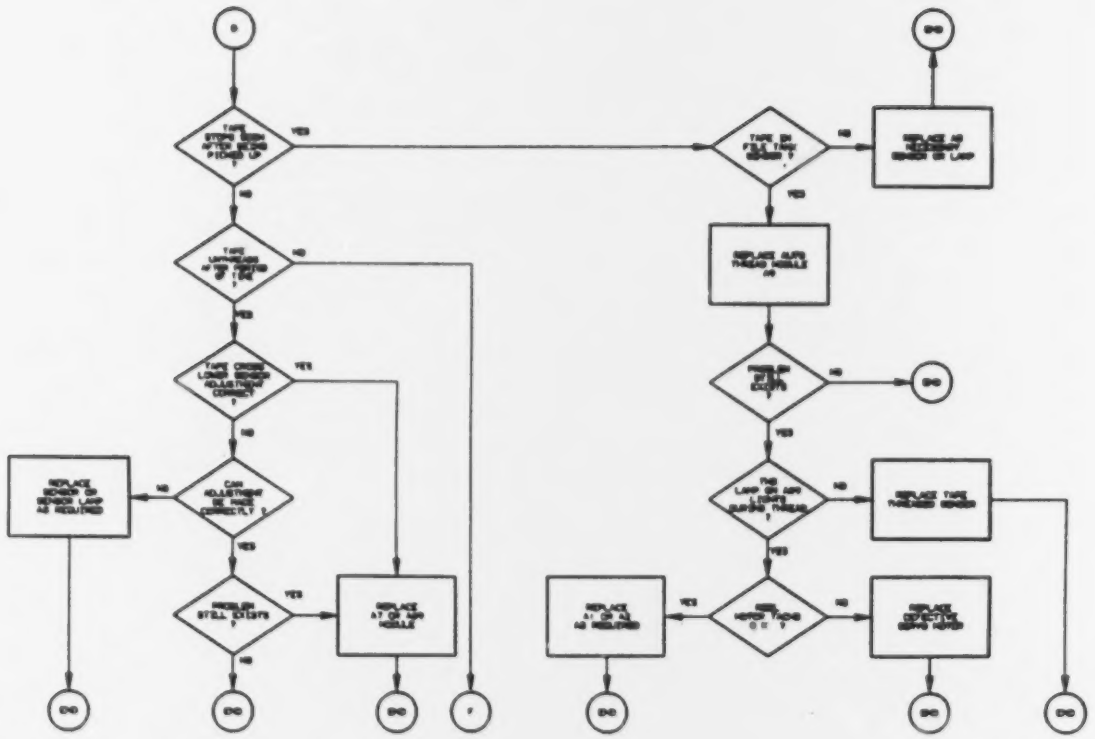

Figure 1. Auto Thread Logic, Fault Isolation Flow Chart

BEST COPY AVAILABLE

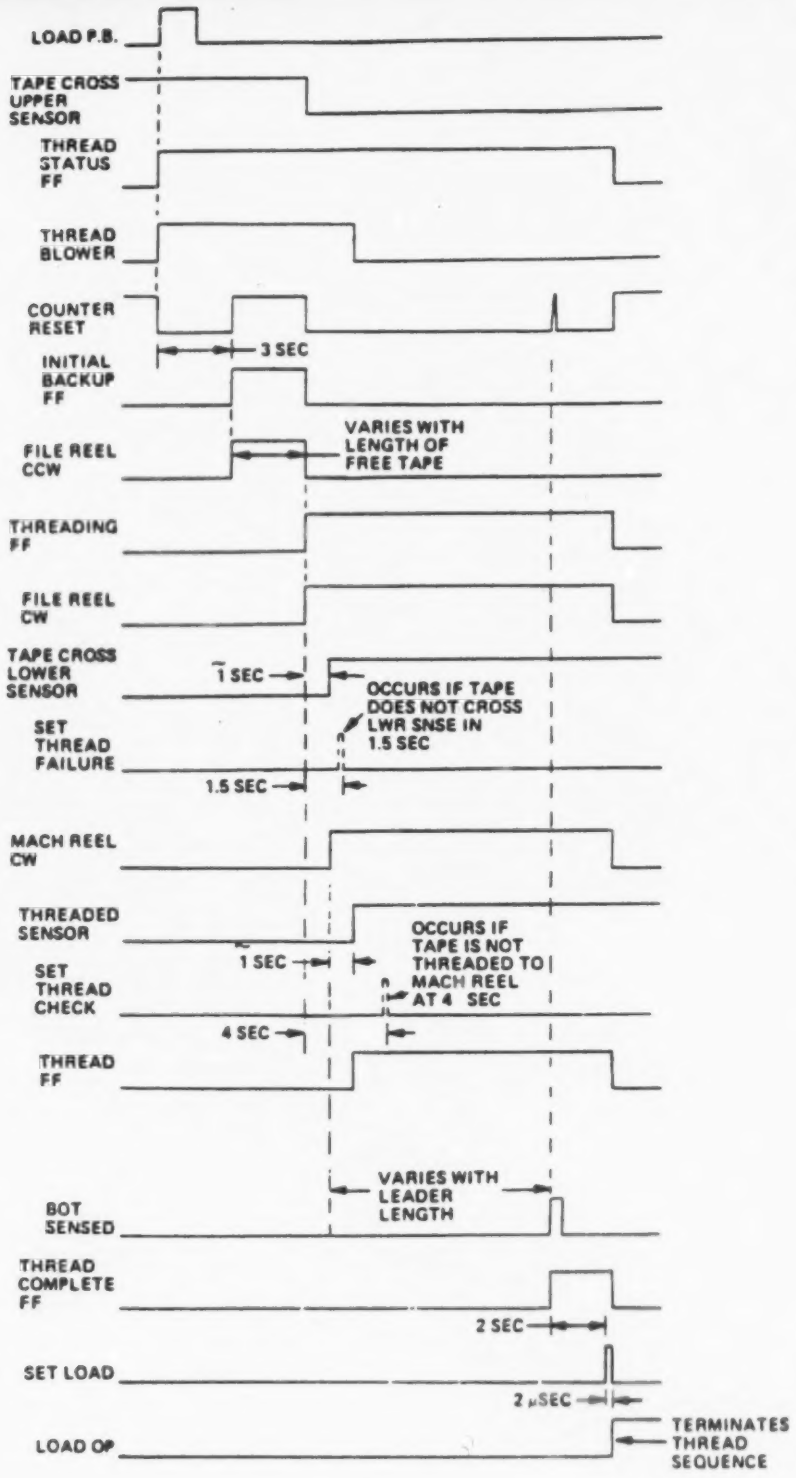

Figure 2. Auto Thread Timing Diagram

BEST COPY AVAILABLE

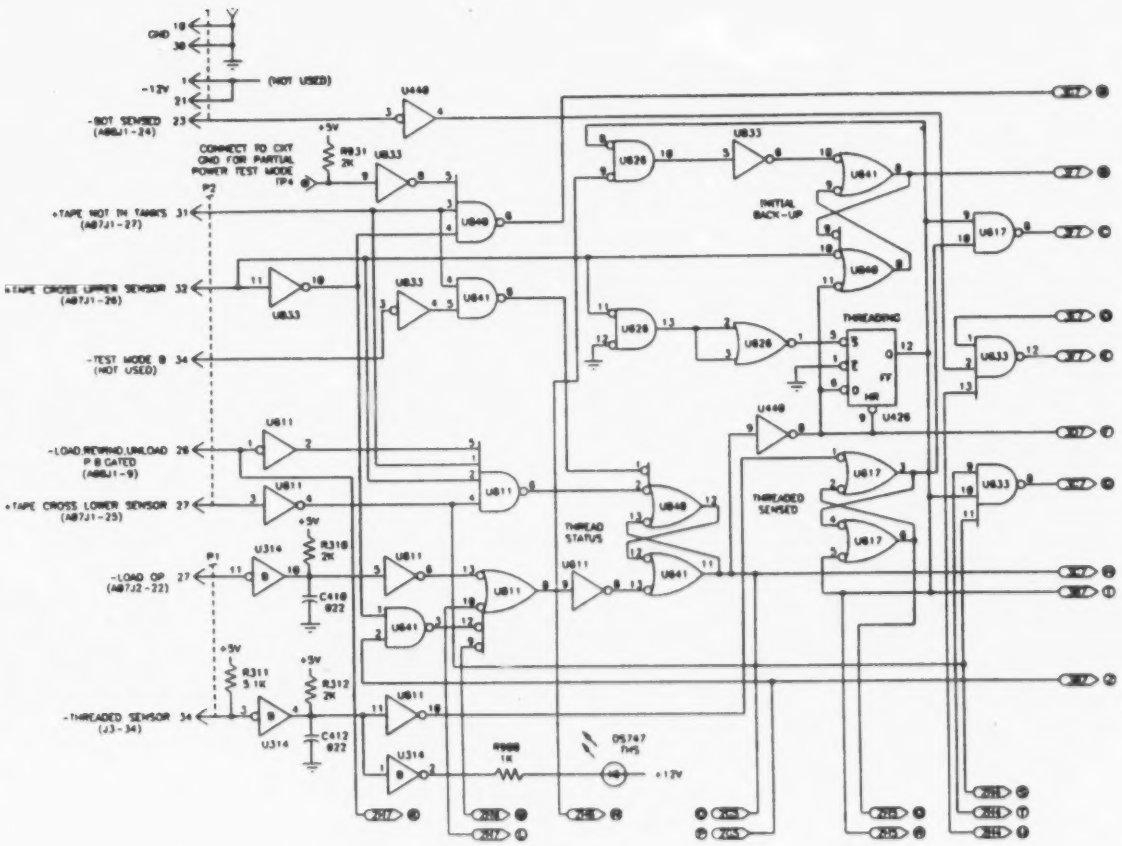

Figure 3. Load/Rewind Functional Block Diagram

BEST COPY AVAILABLE

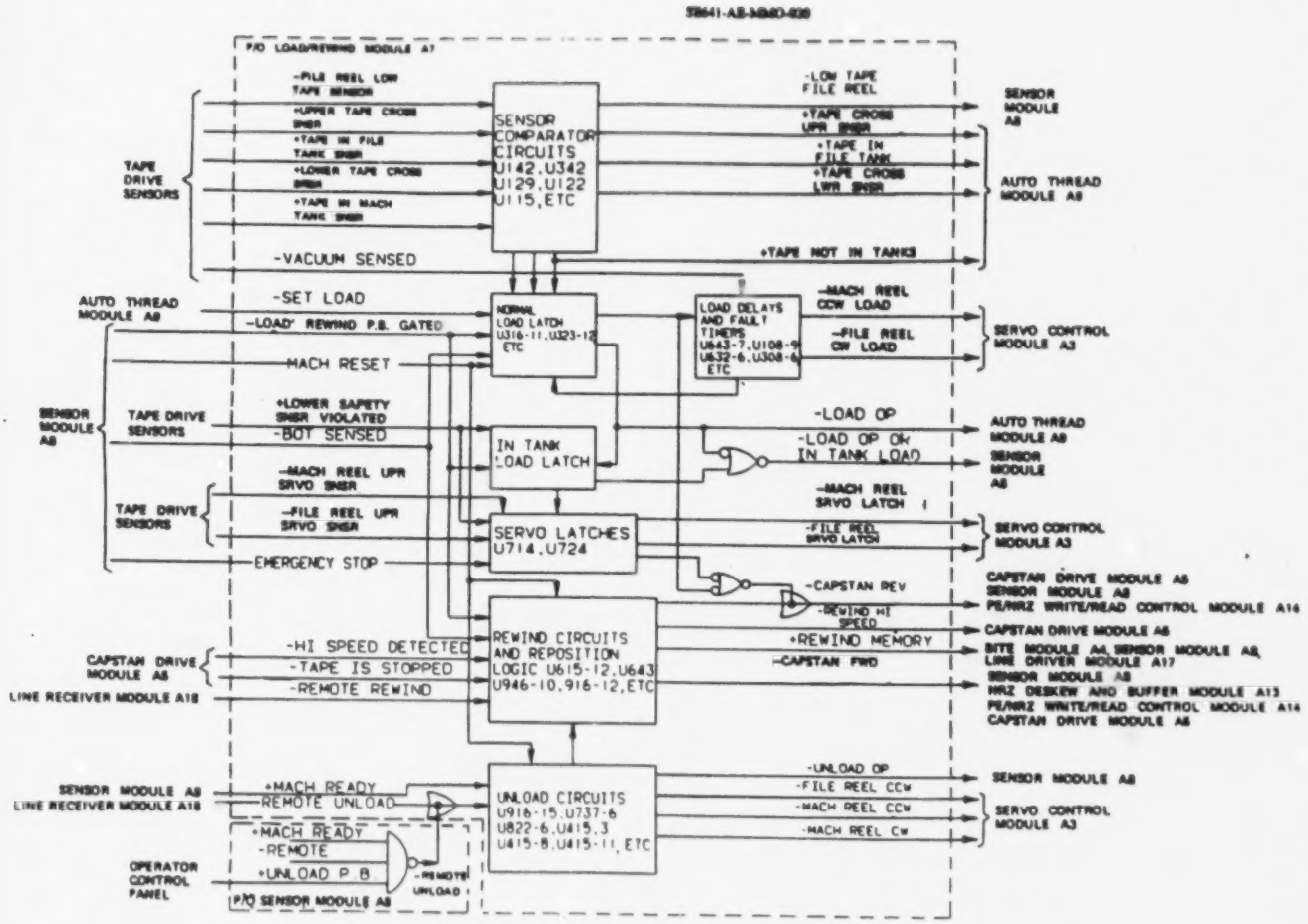

Figure 4. Auto Thread Board, Schematic Diagram

BEST COPY AVAILABLE

We would probably have collapsed the methods in this way if the symbols and processing conventions across the representations were nearly identical, and the differences among them were something like slight changes in scale. In making this choice, we would be claiming that the knowledge to use the four different representations is nearly identical; there would be no diagnostic advantage to testing or instructing separate procedures for using these diagrams. But the appearance of the diagrams makes it clear that the knowledge for using one is quite different than the knowledge for using another, and suggests to us the need to define their uses as separate diagnostic methods.

Another rationale for our interpretation is that the methods have slightly different side effects. The flow chart method primarily dictates action. The other methods provide predictions and explanation. In the present case, the chief could have simply performed the actions recommended by the fault-isolation flow chart. But, he prefers to understand the structure behind the recommendation and pursues other methods of fault isolation in parallel. Notice that the branches of the fault-isolation chart end with a recommended action. If these final actions fail to isolate the fault, the diagnostician would be forced to apply the other methods.

Each of the methods we identify can be further decomposed. The present excerpt provides information to help us decompose "using flowchart", "b" in the plan-goal graph (see Figure 5), which is much more complicated than we expected. One sub-goal for using these flow charts (not illustrated in the excerpt) is simply to locate the correct flow chart ("c" in the plan-goal graph). This is often established by using an index that maps descriptions of problems onto flow chart numbers ("d").

We became aware of this sub-goal for two reasons. First, we have observed trainees who have difficulty using the index for locating the correct flow chart. Second, the session from which the excerpt was drawn includes an episode in which the chief notices after some time that he is using the wrong flow chart. In both cases, the failure to achieve this state of the world (having the correct figure) halts any progress on using the flow charts. We infer that this state must be present in order to use the flow charts properly.

We name goals choosing words that convey states of the world rather than procedures for achieving these states. For these reasons, we avoid goal names that use present tense verbs that suggest action. For example, we named the sub-goal "flow chart applied" to avoid the procedural connotation of the name "apply flow chart". This convention helps to maintain the distinction between goals and plans.

The observed difficulties in locating the correct flow chart illustrate how mistakes inform the task analysis by indicating knowledge requirements that may not be obvious when performance is perfect. In addition, the chief's episode suggests the presence of knowledge for confirming that the appropriate flow chart is in use. Without such knowledge, the chief could not have identified and corrected his error.

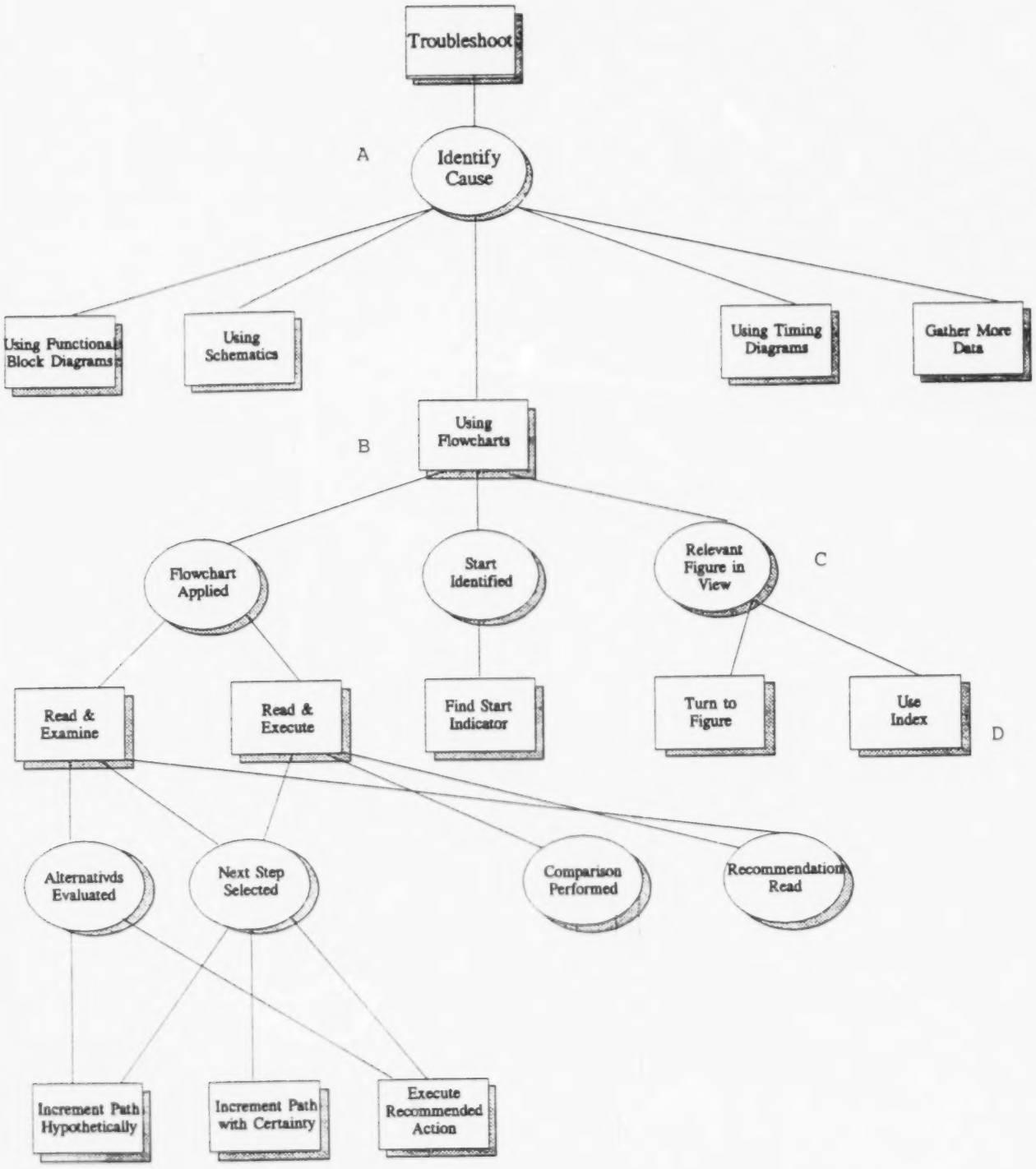

Figure 5. Portions of a plan-goal graph for the computer technician job.

BEST COPY AVAILABLE

Summary

With this example, we illustrated our approach to gathering, analyzing, and representing knowledge from protocol data. The example depicts an approach to identifying the goals and methods of task performance, as well as some common challenges associated with gathering protocol data (e.g., getting subjects to verbalize, managing extraneous influences). Representing this knowledge in a plan-goal graph suits the intermediate level of analyses appropriate to the development of a job knowledge test. By encoding knowledge into the structure of a plan-goal graph, we confirm that each knowledge element is relevant to a goal of task performance.

Thus, we did not formally analyze the protocol data, nor did we implement a computational cognitive model. Rather, we developed an initial plan-goal graph model and refined it through several protocol gathering sessions. This approach very substantially reduces the time, personnel, and other costs that would incur from more formal data analytic methods.

Appendix B:
Item Writing Guidelines for Written Performance Measures

We reviewed the scientific literature to locate guidelines for constructing good tests, with an emphasis on measures of performance. We found that prescriptions from the literature overwhelmingly focus on identifying good questions, or revising poor ones, from a pool of existing questions. Most of this work addresses the use of statistical item indices (e.g., difficulty, discrimination) to assist item revision and to guide selection of items for inclusion in a test. Few guidelines and tools exist for actually constructing good test questions. Much of what does exist has received comparatively little empirical scrutiny (Haladyna & Downing, 1989).

Nevertheless, the advice of experienced test developers is valuable to know. We distilled the following suggestions from the literature, filtered through our perspective on job expertise and performance measurement. A main point of our perspective is that tests should require examinees to use the same information, in the same way, as they would during performance. That is, we emphasize the importance of content (e.g., as opposed to method or format of measurement) to achieving assessment goals of useful diagnostic information and valid predictions of performance. While the emphasis on content is not new to psychological measurement, the task analysis approach described in the report should contribute to improved specifications of what that content should be.

We provide these suggestions for the assistance they may provide to those faced with the task of developing tests and to encourage further research addressing development of test objectives and the construction of written tests and performance measures. The suggestions are organized by a typical sequence of test development activities: specification of test objectives and content, selecting a test format, general rules of item writing, developing item stems, constructing response options, reviewing and revising items, and selecting items for a test. These suggestions were based on the references supplied at the end of this appendix. In particular, we refer the interested reader to Ellis and Wulfeck (1982), Haladyna (1994), Millman and Greene (1989), and Sechrest, Kihlstrom, and Bootzin (1993).

Specifying Test Content

1. Construct questions that require examinees to use the same information, in the same way, as they would during performance.
   A. Identify the content of job expertise.
   B. Specify the test (and training and performance) objectives clearly, and in detail.
   C. Assess only important objectives, essential to learning or performance.
   D. Representatively sample job expertise, across tasks and knowledge content.

In essence, all of the guidelines are elaborations of this first point. For performance measurement, the goal of each guideline is to ensure that the psychological fidelity (Goldstein, Zedeck, & Schneider, 1992) of performance is retained in the test. The most critical contribution to achieving psychological fidelity rests with the adequacy of the task

analyses in capturing job expertise. One approach to representing this expertise in a form useful to test writers is by specifying test content in a process by content matrix (Ellis & Wulfeck, 1982; Millman & Greene, 1989). This assists test writers by providing explicit goals for how the content domain should be sampled. In the body of this report, we refine this approach by elaborating the nature of job expertise in richer detail.

The remaining guidelines provide checkpoints at each stage of transforming the description of expertise from the task analysis into suitable measurements. Test scores reflect a complex set of influences. In addition to examinees' expertise, test scores are affected by differences among examinees' in their comfort and skill in taking exams, reading comprehension and speed, attitudes towards exams, their fatigue, and their motivation to perform well on the particular test at hand. The initial guidelines address identifying and representatively sampling expertise. The remaining guidelines are directed towards reducing or controlling the other, extraneous factors on test scores.

General Rules

2. Use questions that are relevant and fair.
   A. Avoid questions based on opinion.
   B. Avoid misleading or 'tricky' questions.

3. Use simple, clear, and concise language.
   A. Use good grammar and punctuation so items read well.
   B. Minimize the amount of reading necessary for test questions.

In the approach to test development described in Section 3 of the report, we address rule 2 by providing detailed test specifications based on a task analysis of job expertise. The content of each test question is specified by its relation to a particular task and a detailed description of knowledge requirements. Additionally, the relevance of each knowledge requirement is described by the task analysis results in the form of a plan-goal graph.

The goal of rule 3 is to reduce the impact of test-wiseness and reading comprehension and speed on test scores. These threats to test interpretation and validity will be further addressed in other guidelines which follow. For example, one good testing practice is to ensure that the vocabulary and grammar of the text in the test does not exceed that used in important job documents.

Writing Question Stems

The following rules improve the clarity of test questions. This reduces the potential for differing interpretations of questions among examinees.

4. State the stem in question form.
5. State the question in the affirmative whenever possible.
6. Use the 'best answer' format rather than incorrect answer or multiple answer format
7. The problem in the stem should be understandable without reading the options.
8. A longer stem is better than long options.
9. Include in the stem any words that otherwise would be repeated in the option.

Another set of guidelines involves providing item writers with a taxonomy of item types to guide the development of new items. Relevant taxonomies have been proposed based on item format (analogies, sequences, true-false, etc.; Kline, 1986), linguistic transformations (Roid & Haladyna, 1982), and content (Ellis & Wulfeck, 1982; Haladyna, 1994). We selected content as the basis of guidance for item writers. This permits a more direct mapping of item structure to task analysis results by employing the same taxonomy for both activities.

Constructing Response Options

10. Construct distractors based on common errors.
    A. Make all response options plausible to the uninformed.
    B. Resist humor when developing distractors.

Developing options based on typical errors provides the item writer a clear, practical strategy. It also potentially improves the diagnostic value of tests by allowing examinees and examiners the opportunity to review the nature and pattern of incorrect responses. For well constructed tests, this information can be very valuable in pinpointing sources of misconceptions or difficulty.

11. Avoid distractors that can be ruled out on grounds other than domain content.
    A. Make options mutually exclusive and independent.
    B. Extraneous clues in the distractors should be used sparingly (e.g., stereotyped phrases).
    C. Avoid using a key word from the stem in the correct response.
    C. Avoid distractors using 'always' or 'never'.
    D. Do not use "all of the above", "none of the above", etc. as possible answers.

When selected, options such as 'always' provide little diagnostic information. Additionally, such options allow those with superior test taking skills to rule out some options without having to know anything about the domain being tested. Grammatical cues can also tip off which responses are correct or incorrect. The following suggestions address this possibility.

12. Make all options parallel in form.
    A. Keep all options parallel in grammatical form.
    B. Keep the language of options equally professional.
    C. Keep the lengths of response options fairly consistent.

13. Employ a simple, clear, and concise format for item responses.
    A. Consider using only 3, rather than the usual 4 or 5 response options.
    B. List options vertically, not horizontally.
    C. Arrange options in a logical order (e.g., numerical), if one exists.
    D. Identify options with letters instead of numbers.
    E. Emphasize negative words or words of exclusion (e.g., NOT, EXCEPT).

These suggestions reduce the amount of time it takes to read and comprehend questions. The reduced time allows more questions to be added, thus improving the representativeness of the test. Additionally, the potentially irrelevant and biasing factors of reading comprehension and speed may also be reduced.

Reviewing Test Questions

14. Make sure only one option is correct, or one option is clearly the best.
15. Examine, and rule out, alternative interpretations of test scores.
    A. Ensure that each question is relevant and important to the criterion of interest.
    B. Ensure that item difficulty is appropriate to the examinee population (e.g., about 70%).
    C. Ensure that the reading level of the test and the job are the same.
    D. Review question stems for ambiguity and conciseness.
    E. Review response options for plausibility and relevance to the job.
    F. Ensure that the correct response option is varied randomly across positions.
    G. Ensure that questions are independent and options are mutually exclusive.

Revising Test Questions

16. Alter question difficulty by changing the homogeneity of the responses.
17. Alter question difficulty by changing the complexity of the stem.
18. Improve examinee motivation by embedding the question in realistic situations.

Assembling Tests

19. Select questions with a difficulty level of about 70%.
20. Select questions with an item-test or item-criterion correlation higher than .20.
21. Balance the key so that the correct answer appears about equally in each position.
22. Provide clear, simple, and thorough instructions for test administration and test scoring.

References

Aiken, L. R. (1982). Writing multiple choice items to measure higher order educational objectives. *Educational and Psychological Measurement, 42*, 3, 803-806.

Aiken, L. R. (1994). *Psychological testing and assessment* (8th edition). Boston, MA: Allyn and Bacon.

Bejar, I. I., Chaffin, R. & Embretson, S. E. (1991). *Cognitive and psychometric analysis of analogical problem solving*. New York, NY: Springer-Verlag.

Berk, R.A. (Ed., 1984). *A guide to criterion-referenced test construction*. Baltimore, MD: The Johns Hopkins University Press.

Canot, J. A. (1987). Developing multiple choice test items. *Training and Development Journal*, 41, 5, 85-88.

Ellis, J. A., & Wulfeck, W. H. (1982). *Handbook for testing in Navy schools* (NPRDC SR 83-2). San Diego, CA: Navy Personnel Research and Development Center.

Embretson, S. E. (1985). *Test design: Developments in psychology and psychometrics*. Orlando, FL: Academic Press, Inc.

Goldstein, I. L., Zedeck, S., & Schneider, B. (1992). An exploration of the job analysis-content validity process. In N. Schmitt and W. C. Borman (Eds.), *Personnel Selection in Organizations*. San Francisco, CA: Jossey-Bass Publishers. (pp. 3-34).

Gronlund, N. E. (1993). *How to make achievement tests and assessments* (5th edition). Boston, MA: Allyn and Bacon.

Haladyna, T. M. (1994). *Developing and validating multiple-choice test items*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Haladyna, T. M., & Downing, S. M. (1989). A taxonomy of multiple-choice item writing rules. *Applied Measurement in Education, 2*, 1, 37-50.

Haladyna, T. M., & Downing, S. M. (1989). Validity of a taxonomy of multiple-choice item writing rules. *Applied Measurement in Education, 2*, 1, 51-78.

Kline, P. (1986). *A handbook of test construction*. London: Methuen & Co., Ltd.

Millman, J. & Greene, J. (1989). The specification and development of tests of achievement and ability. In R. L. Linn (Ed.) *Educational Measurement* (3rd edition). New York, NY: Macmillan.

Nevo, B. & Jager, R. S. (Eds., 1993). *Educational and psychological testing: The test taker's outlook*. Lewiston, NY: Hogrefe & Huber.

Roid, Gale H. & Haladyna, T. M. (1982). *A technology for test-item writing*. New York: Academic Press.

Sechrest, L., Kihlstrom, J. F., & Bootzin, R. R. (1993). How to develop multiple-choice tests. APS Observer. (pp. 10ff).

Swezey, R. W. (1981). *Individual performance assessment: An approach to criterion-referenced test development*. Reston, VA: Reston Publishing Co., Inc.

Wood, R. (1977). Multiple choice: A state of the art report. *Evaluation in Education, 1*, 3, 191-230.

Distribution List

Dr. Phillip Ackerman
Psychology Department
University of Minnesota
75 E. River Rd.
Minneapolis, MN 55455

Dr. Terry Allard
Code 342CS
Office of Naval Research
800 N. Quincy St.
Arlington, VA 22217-5000

Dr. Nancy S. Anderson
Department of Psychology
University of Maryland
College Park, MD 20742

Dr. Stephen J. Andriole, Chairman
College of Information Studies
Drexel University
Philadelphia, PA 19104

Edward Atkins
13705 Lakewood Ct.
Rockville, MD 20850

Dr. William M. Bart
University of Minnesota
Dept. of Educ. Psychology
330 Burton Hall
178 Pillsbury Dr., S.E.
Minneapolis, MN 55455

Leo Beltracchi
United States Nuclear
Regulatory Commission
Washington DC 20555

Dr. William O. Berry
Director of Life and
Environmental Sciences
AFOSR/NL, N1, Bldg. 410
Bolling AFB, DC 20332-6448

Dr. Thomas G. Bever
Department of Psychology
University of Rochester
River Station
Rochester, NY 14627

Dr. Menucha Birenbaum
Educational Testing
Service
Princeton, NJ 08541

Dr. Werner P. Birke
Personalstammamt der Bundeswehr
Kolner Strasse 262
D-5000 Koeln 90
FEDERAL REPUBLIC OF GERMANY

Dr. Kenneth R. Boff
AL/CFH
Wright-Patterson AFB
OH 45433-6573

Dr. Robert Breaux
Code 252
Naval Training Systems Center
Orlando, FL 32826-3224

Dr. Ann Brown
Graduate School of Education
University of California
EMST-4533 Tolman Hall
Berkeley, CA 94720

Dr. Pat Carpenter
Carnegie-Mellon University
Department of Psychology
Pittsburgh, PA 15213

Dr. Eduardo Cascallar
Educational Testing Service
Rosedale Road
Princeton, NJ 08541

Dr. Michelene Chi
Learning R & D Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15260

Dr. Susan Chipman
Cognitive Science Program
Office of Naval Research
800 North Quincy St.
Arlington, VA 22217-5660

Dr. Raymond E. Christal
UES LAMP Science Advisor
AL/HRMIL
Brooks AFB, TX 78235

Dr. Deborah Claman
National Institute for Aging
Bldg. 31, Room 5C-35
9000 Rockville Pike
Bethesda, MD 20892

Dr. Paul Cobb
Purdue University
Education Building
W. Lafayette, IN 47907

Dr. Rodney Cocking
NIMH, Basic Behavior and
Cognitive Science Research
5600 Fishers Lane, Rm 11C-10
Parklawn Building
Rockville, MD 20857

Director, Life Sciences
Office of Naval Research
Code 114
Arlington, VA 22217-5000

Director, Cognitive and
Neural Sciences, Code 1142
Office of Naval Research
Arlington, VA 22217-5660

Director
Training Systems Department
Code 15A
Navy Personnel R&D Center
San Diego, CA 92152-6800

Library, Code 231
Navy Personnel R&D Center
San Diego, CA 92152-5800

Dr. Magda Colberg
Office of Personnel
Management
1900 E Street, N.W.
Washington, DC 20415-0001

Commanding Officer
Naval Research Laboratory
Code 4827
Washington, DC 20375-5000

Dr. Albert T. Corbett
Department of Psychology
Carnegie-Mellon University
Pittsburgh, PA 15213

Dr. Kenneth B. Cross
Anacapa Sciences, Inc.
P.O. Box 519
Santa Barbara, CA 93102

Dr. Charles E. Davis
Educational Testing Service
Mail Stop 22-T
Princeton, NJ 08541

Dr. Geory Delacote
Exploratorium
3601 Lyon Street
San Francisco, CA
94123

Chief of Personnel
Testing Division
Defense Manpower Data Ctr
99 Pacific Street Suite 155A
Monterey CA 93940-2453

Dr. Sharon Derry
Florida State University
Department of Psychology
Tallahassee, FL 32306

Defense Technical
Information Center
DTIC/DDA-2
Cameron Station, Bldg 5
Alexandria, VA 22314

David DuBois
Psychological Systems & Research, Inc.
1975 Willow Ridge Circle
Kent, OH 44240

Dr. Richard Duran
Graduate School of Education
University of California
Santa Barbara, CA 93106

Dr. Nancy Eldredge
College of Education
Division of Special Education
The University of Arizona
Tucson, AZ 85721

Dr. John Ellis
Navy Personnel R&D Center
Code 15
San Diego, CA 92152-6800

ERIC Facility-Acquisitions
1301 Piccard Drive, Suite 300
Rockville, MD 20850-4305

Dr. K. Anders Ericsson
University of Colorado
Department of Psychology
Campus Box 345
Boulder, CO 80309-0345

Dr. Martha Evens
Dept. of Computer Science
Illinois Institute of Technology
10 West 31st Street
Chicago, IL 60616

Dr. Lorraine D. Eyde
US Office of Personnel Management
Office of Personnel Research and
Development
1900 E St., NW
Washington, DC 20415

Dr. Franco Faina
Direttore Generale LEVADIFE
Piazzale K. Adenauer, 3
00144 ROMA EUR
ITALY

Dr. Beatrice J. Farr
Army Research Institute
PERI-IC
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. Marshall J. Farr
Farr-Sight Co.
2520 North Vernon Street
Arlington, VA 22207

Dr. Lawrence T. Frase
Executive Director
Division of Cognitive and
Instructional Science
Educational Testing Service
Princeton, NJ 08541

Dr. Norman Frederiksen
Educational Testing Service
(05-R)
Princeton, NJ 08541

Chair, Department of
Computer Science
George Mason University
Fairfax, VA 22030

Dr. Alan S. Gevins
EEG Systems Laboratory
51 Federal Street, Suite 401
San Francisco, CA 94107

Dr. Helen Gigley
Office of Navel Research
800 N. Quincy Street
Arlington, VA 22217-5660

Dr. Herbert Ginsburg
Box 184
Teachers College
Columbia University
525 West 121st Street
New York, NY 10027

Dr. Drew Gitomer
Educational Testing Service
Princeton, NJ 08541

Dr. Robert Glaser
Learning Research
& Development Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15260

Dr. Paul E. Gold
University of Virginia
Department of Psychology
Charlottesville, VA 22903

Dr. Susan R. Goldman
Peabody College, Box 45
Vanderbilt University
Nashville, TN 37203

Dr. Timothy Goldsmith
Department of Psychology
University of New Mexico
Albuquerque, NM 87131

Deborah F. Goodman
University of Nebraska at Omaha
CBA 308L
60th & Dodge Streets
Omaha, NE 68182-0459

Dr. Sherrie Gott
AFHRL/MOMJ
Brooks AFB, TX 78235-5601

Dr. Wayne Gray
Graduate School of Education
Fordham University
113 West 60th Street
New York, NY 10023

Dr. Bert Green
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore, MD 21218

Prof. Lutz F. Hornke
Institut fur Psychologie
RWTH Aachen
Jaegerstrasse 17/19
D-5100 Aachen
WEST GERMANY

Ms. Julia S. Hough
Cambridge University Press
40 West 20th Street
New York, NY 10011

Dr. William Howell
Executive Director for Science
750 First Street
Washington DC 20002-4242

Dr. Eva Hudlicka
BBN Laboratories
10 Moulton Street
Cambridge, MA 02238

Dr. Earl Hunt
Dept. of Psychology, NI-25
University of Washington
Seattle, WA 98195

Dr. Martin J. Ippel
Center for the Study of
Education and Instruction
Leiden University
P. O. Box 9555
2300 RB Leiden
THE NETHERLANDS

Dr. Robert Jannarone
Elec. and Computer Eng. Dept.
University of South Carolina
Columbia, SC 29208

Dr. Edgar M. Johnson
Technical Director
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333-5600

Dr. Peder Johnson
Department of Psychology
University of New Mexico
Albuquerque, NM 87131

Dr. John Jonides
Department of Psychology
University of Michigan
Ann Arbor, MI 48104

Dr. Marcel Just
Carnegie-Mellon University
Department of Psychology
Schenley Park
Pittsburgh, PA 15213

Dr. Michael Kaplan
Office of Basic Research
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333-5600

Dr. Sung-Ho Kim
Educational Testing Service
Princeton, NJ 08541

Dr. Stephen Kosslyn
Harvard University
1236 William James Hall
33 Kirkland St.
Cambridge, MA 02138

Dr. Kenneth Kotovsky
Department of Psychology

BEST COPY AVAILABLE

Carnegie-Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213

Dr. Richard J. Koubek
School of Industrial
Engineering
Grissom Hall
Purdue University
West Lafayette, IN 47907

Dr. Art Kramer
Univ. of Illinois at U-C
Beckman Institute
405 N. Mathews Avenue
Urbana, IL 61801

Dr. Michael Kuperstein
Symbus Technology
325 Harvard Street
Suite 211
Brookline, MA 02146

Dr. Patrick Kyllonen
AFHRL/MOEL
Brooks AFB, TX 78235

Dr. Marcy Lansman
University of North Carolina
Dept. of Computer Science
CB #3175
Chapel Hill, NC 27599

Dr. Robert W. Lawler
Matthews 118
Purdue University
West Lafayette, IN 47907

Dr. Michael Levine
Educational Psychology
210 Education Bldg.
1310 South Sixth Street
University of IL at
Urbana-Champaign
Champaign, IL 61820-6990

Dr. Robert Levin
The Center for Human Function & Work
1526 Spruce Street
Boulder, CO 80302

Logicon Inc. (Attn: Library)
Tactical and Training Systems
Division
P.O. Box 85158
San Diego, CA 92138-5158

Prof. David F. Lohman
College of Education
University of Iowa
Iowa City, IA
52242

Dr. Scott Makeig
Naval Health Research
Center, Bldg. 331
San Diego, CA

Vern M. Malec

NPRDC, Code 142
San Diego, CA 92152-6800

Dr. Sandra P. Marshall
Dept. of Psychology
San Diego State University
San Diego, CA 92182

Dr. Elizabeth Martin
AL/HRA, Stop 44
Williams AFB
AZ 85240

Dr. Joseph McLachlan
Navy Personnel Research
and Development Center
Code 14
San Diego, CA 92152-6800

Dr. Vittorio Midoro
CNR-Istituto Tecnologie Didattiche
Via All'Opera Pia 11
GENOVA-ITALIA 16145

Dr. Robert Mislevy
Educational Testing Service
Rosedale Rd.
Princeton, NJ 08541

Dr. Allen Munro
Behavioral Technology
Laboratories - USC
250 N. Harbor Dr., Suite 309
Redondo Beach, CA 90277

Academic Progs. & Research Branch
Naval Technical Training Command
Code N-62
NAS Memphis (75)
Millington, TN 30854

Director
Training Systems Department
NPRDC (Code 14)
San Diego, CA 92152-6800

Library, NPRDC
Code 041
San Diego, CA 92152-6800

Dr. Harold F. O'Neil, Jr.
School of Education - WPH 600
Department of Educational
Psychology & Technology
University of Southern California
Los Angeles, CA 90089-0031

Office of Naval Research,
Code 1142CS
800 N. Quincy Street
Arlington, VA 22217-5000
(6 Copies)

Dr. Judith Orasanu
Mail Stop 239-1
NASA Ames Research Center
Moffett Field, CA 94035

Dr. Everett Palmer

Mail Stop 262-4
NASA-Ames Research Center
Moffett Field, CA 94035

Dr. Roy Pea
Institute for the
Learning Sciences
Northwestern University
1890 Maple Avenue
Evanston, IL 60201

G. Pelsmakers
Rue Fritz Toussaint 47
Gendarmerie RSP
1050 Bruxelles
BELGIUM

Dr. Ray S. Perez
ARI (PERI-II)
5001 Eisenhower Avenue
Alexandria, VA 22333

C.V. (MD) Dr. Antonio Peri
Captain ITNMC
Maripers U.D.G. 3' Sez
MINISTERO DIFESA - MARINA
00100 ROMA - ITALY

Dr. Nancy N. Perry
Naval Education and Training
Code 047
Building 2435
Pensacola, FL 32509-5000

CDR Frank C. Petho
Naval Postgraduate
School
Code OR/PE
Monterey, CA 93943

Dept. of Administrative Sciences
Code 54
Naval Postgraduate School
Monterey, CA 93943-5026

Dr. Peter Pirolli
School of Education
University of California
Berkeley, CA 94720

Dr. Martha Polson
Department of Psychology
University of Colorado
Boulder, CO 80309-0344

Dr. Peter Polson
University of Colorado
Department of Psychology
Boulder, CO 80309-0344

Dr. Joseph Psotka
ATTN: PERI-IC
Army Research Institute
5001 Eisenhower Ave.
Alexandria, VA 22333-5600

Psyc Info - CD and M
American Psychological Assoc.
1200 Uhle Street

Arlington, VA 22201

Dr. Malcolm Ree
AC/HRMA
7909 Lindbergh Dr.
Brooks AFB, TX 78235

Dr. J. Wesley Regian
Armstrong Laboratory
AFHRL/IDI
Brooks AFB, TX 78235-5000

Dr. Brian Reiser
Institute for the Learning Sciences
Northwestern University
1890 Maple Avenue
Evanston, IL 60201-3142

Dr. Lauren Resnick
Learning R & D Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15213

Dr. Linda G. Roberts
Science, Education, and
Transportation Program
Office of Technology Assessment
Congress of the United States
Washington, DC 20510

Dr. Salim Roukos
IBM Corporation
T. J. Watson Research Center
PO Box 218
Yorktown Heights, NY 10598

Dr. Eduardo Salas
Human Factors Division (Code 262)
12350 Research Parkway
Naval Training Systems Center
Orlando, FL 32826-3224

Dr. Fumiko Samejima
Department of Psychology
University of Tennessee
310B Austin Peay Bldg.
Knoxville, TN 37966-0900

Mr. Drew Sands
NPRDC Code 62
San Diego, CA 92152-6800

Dr. Walter Schneider
Learning R&D Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15260

Dr. Myrna F. Schwartz
Director
Neuropsychology Research Lab
Moss Rehabilitation Hospital
1200 West Tabor Road
Philadelphia, PA 19141

Dr. Robert J. Seidel
US Army Research Institute
5001 Eisenhower Ave.
Alexandria, VA 22333

Dr. Michael G. Shafto
NASA Ames Research Ctr.
Mail Stop 262-1
Moffett Field, CA 94035-1000

Dr. Tracy Shors
Dept. of Psychology
Princeton Univ.
Green Hall
Princeton, NJ 08544

Dr. Zita M. Simutis
Director, Manpower & Personnel
Research Laboratory
US Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333-5600

Dr. Derek Sleeman
Computing Science Department
The University
Aberdeen AB9 2FX
Scotland
UNITED KINGDOM

Dr. Robert Smillie
Naval Ocean Systems Center
Code 443
San Diego, CA 92152-5000

Dr. Richard E. Snow
School of Education
Stanford University
Stanford, CA 94305

Dr. Bruce D. Steinberg
Curry College
Milton, MA 02186

Dr. Michael J. Tarr
Dept. of Psychology
Yale University
PO Box 208205 Yale Station
New Haven, CT 06520-8205

Dr. Kikumi Tatsuoka
Educational Testing Service
Mail Stop 03-T
Princeton, NJ 08541

Chair, Department of Psychology
University of Maryland,
Baltimore County
Baltimore, MD 21228

Dr. Kurt VanLehn
Learning Research
& Development Ctr.
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15260

Dr. Frank L. Vicino
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Jerry Vogt
Department of Psychology
St. Norbert College
De Pere, WI 54115-2099

Dr. Jacques Voneche
University of Geneva
Department of Psychology
Geneva
SWITZERLAND 1204

Dr. Barbara White
School of Education
Tolman Hall, EMST
University of California
Berkeley, CA 94720

Dr. David Wiley
School of Education
and Social Policy
Northwestern University
Evanston, IL 60208

Dr. David C. Wilkins
University of Illinois
Department of Computer Science
405 North Mathews Avenue
Urbana, IL 61801

Dr. Mark Wilson
Graduate School of
Education
University of California,
Berkeley
Berkeley, CA 94720

Dr. Robert A. Wisher
U.S. Army Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333-5600

Dr. Merlin C. Wittrock
Graduate School of Education
Univ. of Calif., Los Angeles
Los Angeles, CA 90024

Dr. Kentaro Yamamoto
03-0T
Educational Testing Service
Rosedale Road
Princeton, NJ 08541

Dr. Masoud Yazdani
Dept. of Computer Science
University of Exeter
Prince of Wales Road
Exeter EX44PT
ENGLAND

Frank R. Yekovich
Dept. of Education
Catholic University
Washington, DC 20064

Dr. Joseph L. Young
National Science Foundation
Room 320
1800 G Street, N.W.
Washington, DC 20550